Deep Networks Provably Classify Data on Curves
Supplemental

Tingran Wang
Columbia University
tw2579@columbia.edu

Sam Buchanan
Columbia University
s.buchanan@columbia.edu

Dar Gilboa
Harvard University
dar_gilboa@fas.harvard.edu

John Wright
Columbia University
jw2966@columbia.edu

Contents

A Details of Figures
A.1 Figure 2 .................................................. 2
A.2 Figure 3 .................................................. 2
A.3 Kernel Decay Scale and Trainability of Realising Networks: Empirical Evidence . 4

B Notation .................................................. 5

C Key Definitions
C.1 Problem Formulation ..................................... 6
C.2 Geometric Properties .................................... 8
C.3 Subspace of Smooth Functions and Kernel Derivatives ............................. 10

D Main Results ............................................ 10

E Proof for the Certificate Problem
E.1 Invertibility Over a Subspace of Smooth Functions ................................. 13
E.2 Certificates for the DC-Subtracted Kernel ............................................ 29
E.3 Certificates with Density and DC ....................................................... 51

F Bounds for the Skeleton Function $\psi$
F.1 Sharp Lower Bound for the Iterated Angle Evolution Function ................ 60
F.2 Sharp Lower Bound for $\psi$ ....................................................... 68
F.3 Nearly-Matching Upper Bound .................................................... 70
F.4 Higher Order Derivatives of $\psi$ ................................................... 74
F.5 Additional Proofs for Some Bounds ............................................. 78

We give full implementation details for this figure here, mixed with conceptual ideas that underlie the
implementation. The manifolds \( M_+ \) and \( M_- \) are defined by parametric equations \( x_+ : [0, 1] \to S^2 \)
and \( x_- : [0, 1] \to S^2 \); it is not practical to obtain unit-speed parameterizations of general curves,
so we also have parametric equations for their derivatives \( \dot{x}_\sigma : [0, 1] \to \mathbb{R}^2 \). These are important in

\[
\begin{align*}
(x_{-,1}(t)) &= \left( \frac{\cos(4t)}{\delta} \cos(t) (\sin(4t) + 1 + \delta) + \sin \left( \frac{\pi}{8} \right) \sin(t) (\sin(t) + 1 + \delta) \\
(x_{-,2}(t)) &= \left( -\sin \left( \frac{\pi}{8} \right) \cos(t) (\sin(4t) + 1 + \delta) + \cos \left( \frac{\pi}{8} \right) \sin(t) (\sin(t) + 1 + \delta) \\
(x_{-,3}(t)) &= 4 \sin(t) \\
(x_{+,1}(t)) &= 4 \cos(t) - 1 \\
(x_{+,2}(t)) &= 0 \\
(x_{+,3}(t)) &= 0
\end{align*}
\]

where \( \delta \) sets the separation between the manifolds and is set here to \( \delta = 0.05 \). We then rescale both
curves by a factor \( 0.1 \): the scale of the curves is chosen such that the curvature of the sphere has a
negligible effect on the curvature of the manifolds (since the chart mapping we use here distorts the
curves more nearer to the boundary of the unit disk \( \{(u, v, w) \mid u^2 + v^2 + w^2 \leq 1\} \)).

From here, we use an “unfolding” process to obtain the blue curves in the other three panels from
\( x_- \). To do this, points where \( \left| \frac{dx}{dt} \right| = \left| \frac{dx_-}{dt} \right| \) are found numerically. There are 8 such points in
total, and parts of the curve between pairs of these points are reflected across the line defined by such
a pair in the \( (x_2, x_3) \) plane. This can be done for any number of pairs between 1 and 4, generating
the curves shown. This procedure ensures that aside from the set of 8 points, the curvature at every
point along the curve is preserved and there is no discontinuity in the first derivative, while making
the geometries loop back to the common center point more. For an additional visualization of the
geometry, see Figure 1.  

Given these geometries, in order to compute the certificate norm for the experiment in the top-right
panel, we evaluate the resulting curves at 200 points each, chosen by picking equally spaced points in
\([0, 2\pi]\) and evaluating the parametric equations. The certificate itself is evaluated numerically as in
Appendix A.2.

**Rotated MNIST digits.** We rotate an MNIST image around its center by \( i \* \pi/100 \) for integer \( i \)
between 0 and 199. We then apply t-SNE [2] using the scikit-learn package with perplexity 20 to
generate the embeddings.

### A.2 Figure 3

We give full implementation details for this figure here, mixed with conceptual ideas that underlie the
implementation. The manifolds \( M_+ \) and \( M_- \) are defined by parametric equations \( x_+ : [0, 1] \to S^2 \)
and \( x_- : [0, 1] \to S^2 \); it is not practical to obtain unit-speed parameterizations of general curves,
so we also have parametric equations for their derivatives \( \dot{x}_\sigma : [0, 1] \to \mathbb{R}^2 \). These are important in

1. Although this adds a minor confounding effect to our experiments with certificate norm in the top-right
   panel, it is suppressed by setting the scale sufficiently small, and it can be removed in principle by using an
   isometric chart for the upper hemisphere instead of the map given above.

2. For a three-dimensional interactive visualization, see https://colab.research.google.com/drive/
   1xmpYeLk60GdxXOk3Et_apAniEB9fARRv?usp=sharing.
Figure 1: The two curve geometry described in Appendix A.1. The different choices of $M_-$ that lead to different $\circledast$-number are overlapping. The legend indicates the $\circledast$-number of the two curves problem obtained by considering the same $M_+$ but a different $M_-$ as indicated by the color.

our setting since for non-unit-speed curves, the chain rule gives for the integral of a function (say) $f: M_+ \to \mathbb{R}$:

$$\int_{M_+} f(x) \, dx = \int_{[0,1]} (f \circ x_+(t)) \|\dot{x}_+(t)\|_2 \, dt.$$  

In particular, in our experiments, we want to work with a uniform density $\rho = (\rho_+, \rho_-)$ on the manifolds, where the classes are balanced. To achieve this, use the previous equation to get that we require

$$1 = \int_{M_+} \rho_+(x) \, dx + \int_{M_-} \rho_-(x) \, dx$$

$$= \int_{M_+} (\rho_+ \circ x_+)(t) \|\dot{x}_+(t)\|_2 \, dt + \int_{M_-} (\rho_- \circ x_-)(t) \|\dot{x}_-(t)\|_2 \, dt.$$

A uniform density on $M$ is not a constant value—rather, it is characterized by being translation-invariant. It follows that $\rho_\sigma$ should be defined by

$$\rho_\sigma \circ x_\sigma(t) = \frac{1}{2 \|\dot{x}_\sigma(t)\|_2}.$$  

For the experiment, we solve a discretization of the certificate problem, for which the above ideas will be useful. Consider $\Theta$ in (3.1) for a fixed depth $L$ (and $n = 2$, since width is essentially irrelevant here). By the above discussion, the certificate problem in this setting is to solve for the certificate $g = (g_+, g_-)$

$$f_* = \frac{1}{2} \left( \int_{[0,1]} \Theta(\cdot, x_+(t))g_+ \circ x_+(t) \, dt + \int_{[0,1]} \Theta(\cdot, x_-(t))g_- \circ x_-(t) \, dt \right).$$  

Here, we have eliminated the initial random neural network output $f_{\theta_0}$ from the RHS. Aside from making computation easier, this is motivated by fact that the network output is approximately piecewise constant for large depth $L$, and we therefore expect it not to play much of a role here.

Let $M \in \mathbb{N}$ denote the discretization size. Then a finite-dimensional approximation of the previous integral equation is given by the linear system

$$f_* \circ x_\sigma(t_i) = \frac{1}{2M} \left( \sum_{j=1}^{M} \Theta(x_\sigma(t_i), x_+(t_j))g_+ \circ x_+(t_j) + \sum_{j=1}^{M} \Theta(x_\sigma(t_i), x_-(t_j))g_- \circ x_-(t_j) \right)$$

for all $i \in [M]$ and $\sigma \in \{\pm 1\}$, and where $t_i = (i-\frac{1}{M})$. Of course, $f_* \circ x_\sigma(t) = \sigma$, so the equation simplifies further, and because the kernel $\Theta$ and this target $f_*$ are smooth, there is a convergence
of the data in this linear system in a precise sense to the data in the original integral equation as \( M \to \infty \). In particular, define a matrix \( T^+ \) by \( T^+_{ij} = \Theta(x_+(t_i), x_+(t_j)) \), define a matrix \( T^- \) by \( T^-_{ij} = \Theta(x_-(t_i), x_-(t_j)) \), and define a matrix \( T^\pm \) by \( T^\pm_{ij} = \Theta(x_+(t_i), x_-(t_j)) \), all of size \( M \times M \). Then the \( 2M \times 2M \) linear system

\[
\begin{bmatrix}
1 & -1
\end{bmatrix} = \frac{1}{2M} \begin{bmatrix}
T^+ & T^\pm
\end{bmatrix} \begin{bmatrix}
g_+
g_-
\end{bmatrix}
\]

(A.2)

is equivalent to the discretization in (A.1). We implement and solve the system in (A.2) using the definitions we have given above, using the pseudoinverse of the \( 2M \times 2M \) matrix appearing in this expression to obtain \( [g_+, g_-]^\ast \), and plot the results in Figure 3, in particular interpreting \( (g_\sigma) \) as the sampled point \( g_\sigma \circ x_\sigma(t_i) \) as in (A.1) when we plot in the left panel of Figure 3. Evidently, it would be immediate to modify the experiment to replace the LHS of (A.1) by the error \( f_\theta_0 - f_\ast \): the same protocol given above would work, but there would be an element of randomness added to the experiments.

Specifically, in Figure 3 we set \( M = 900 \). When plotting the solution to (A.2), i.e. the vector \( [g_+, g_-]^\ast \), we moreover scale the vector by a factor of 0.3 to facilitate visualization.

### A.3 Kernel Decay Scale and Trainability of Realising Networks: Empirical Evidence

![Graph showing the decay properties of the NTK are predictive of trainability on a toy dataset.](image)

**Figure 2:** The decay properties of the NTK are predictive of trainability on a toy dataset. We plot the log training error of networks of varying depth that are trained to classify two curves with varying separation. The insets show a projection of the geometry onto the plane for separation values 0.3 and 0.9. For each depth \( L \), the characteristic decay scale of the DC-subtracted NTK \( (\psi^\circ) \) is computed numerically and plotted in green. We find that small training loss is only achievable if the decay scale of the kernel is small compared to the inter-manifold distance, hence the decay scale is predictive of trainability.

One of the main insights into the manifold classification problem that is utilized to obtain Theorem 3.2 is that (roughly speaking) the depth of a fully-connected network controls the decay properties of the
network’s NTK, and that fitting can be guaranteed once the decay occurs on a spatial scale that is small relative to certain geometric properties of the data. Here we provide empirical evidence that this phenomenon holds beyond the regime in which our main theorems hold, and in fact is relevant for networks of moderate width and depth as well.

We draw 400 samples each from a uniform distribution over a union of two curves that are related by a rotation by a geodesic angle that is varied from 0.2 to 1.0 in increments of 0.1. The curves are not linearly separable even for large angle (see insets in Fig. 2). These curves are embedded in $S^{n-1}$ for $n_0 = 128$ and subjected to a rotation drawn uniformly from the Haar measure. We then train a fully-connected network to classify the curves using $\ell^1$ loss. The network has width $n = 256$ and we vary the depth from $L = 2$ to $L = 10$, and train using full-batch gradient descent for $10^3$ iterations with learning rate $\tau = 1/(4nL)$ (so that the total effective “training time” is independent of depth). We plot the log training error after training as a function of depth and the inter-manifold distance. For each depth $L$, we estimate an effective “decay scale” of the DC-subtracted skeleton $\psi^p$ by determining the point $s^\star$ such that $\psi^p(s^\star) = \psi^p(0)/2$.

The results are presented in Fig. 2. We observe that the network convergences to small training loss only when the depth is large comparable to the inverse of the manifold separation. As the depth represents the decay rate of the NTK, this indicates that a deeper network generates a localized NTK, allowing faster decay of the training error and making the classification problem easy. Notice that since the geometry of the dataset and network architecture do not satisfy all the assumptions of Theorem 3.2, the experiment provides evidence that the underlying phenomena regarding the role of the depth hold in greater generality. This preliminary result also suggests that the connection between the network architecture and the data geometry, as expressed through the decay properties of the NTK, can have a dramatic effect on the training process even for fully-connected networks.

### B Notation

We use bold lowercase $x$ for vectors and uppercase $A$ for matrices and operators. We generally use non-bold notation to represent scalars and scalar-valued functions. $\mathbb{R}, \mathbb{C}, \mathbb{Z}$ are used for the real numbers, complex numbers and integers, respectively. $\mathbb{N}_0$ represents non-negative integers, and $\mathbb{N}$ represents the natural numbers. $\mathbb{R}^n$ represents $n$-dimensional Euclidean space, $\mathbb{C}^n$ represents the space of complex $n$-tuples (as a $n$-dimensional vector space over $\mathbb{C}$) and $\mathbb{S}^{n-1} \subset \mathbb{R}^n$ represents the $n-1$ dimensional sphere centered at zero with unit radius. For a complex number $z = x + iy$ (or a complex-valued function), $|z| = \sqrt{x^2 + y^2}$ denotes the complex modulus, and $z = x - iy$ denotes the complex conjugate. For $x, y \in \mathbb{C}^n$, we denote $\|x\|_p = (\sum_{i=1}^n |x_i|^p)^{1/p}$ for any $p \in \mathbb{R}$ as the $p$-norm and $\langle x, y \rangle = \sum_{i=1}^n x_i y_i$ as the standard (second-argument-linear) inner product. We use $x^\dagger$ and $A^*$ to represent the conjugate transpose of vectors or matrices of complex numbers (so e.g. $x^\dagger y = \langle x, y \rangle$).

We use $P_S$ to represent the orthogonal projection operator onto a closed subspace $S$ of a normed vector space (typically a Hilbert space).

For a Borel measure space $(X, \mu)$ and any measurable function $f : X \to \mathbb{C}$, we use $\|f\|_{L_p^\mu} = (\int_X |f(x)|^p d\mu(x))^{1/p}$ to represent the $L^p$ norm of $f$ for $0 < p < \infty$. We omit the measure from the notation when it is clear from context. For $p = \infty$, we use $\|f\|_{L_{\infty}^\mu} = \inf \{ C \geq 0 \mid \|f(x)\| \leq C \text{ for } \mu\text{-almost every } x \}$ to represent its essential supremum. We denote the $L^p$ space of $(X, \mu)$ by $L_p^\mu(X)$ (or simply $L_p^\mu$ when the space is clear from context), which is formed by all complex-valued measurable functions with finite $L_p^\mu$ norm. For another space $(Y, \nu)$ and a (linear) operator $T : L_p^\mu(X) \to L_q^\nu(Y)$, we represent its $L_p^\mu \to L_q^\nu$ operator norm as $\|T\|_{L_p^\mu \to L_q^\nu} = \sup_{\|f\|_{L_p^\mu} = 1} \|T[f]\|_{L_q^\nu}$.

When $X = Y$, $\mu = \nu$, and $p = q = 2$ (and $(X, \mu)$ is sufficiently regular), we have a Hilbert space; we write $\langle f, g \rangle_{L_2^\mu} = \int_X f(x) \overline{g(x)} d\mu(x)$ for the inner product, and $T^*$ to denote the associated adjoint of an operator $T$ (so e.g. $f^* = (f, \cdot)$ denotes the corresponding dual element of a function $f$). We use $\text{Id} : L_p^\mu(X) \to L_p^\mu(X)$ to denote the identity operator, i.e. $\text{Id}[f] = f$ for every $f \in L_p^\mu$. For $S \subset X$, we use $\mathbb{1}_S$ to represent the indicator function $\mathbb{1}_S(x) = 1, \forall x \in S$ and 0 otherwise; we will write $\mathbb{1}_X$ to denote $\mathbb{1}_Y$. For a map $\varphi : X \to X$ and $i \in N$, we use $\varphi[i]$ to denote its $i$-th fold iterated composition of itself, i.e. $\varphi[i](x) = \varphi(\varphi^{(i-1)}(x))$. For $i \in N$, $f^{(i)}$ is normally used to represent a function of a real variable $f$’s $i$-th order derivatives. For example, when
the space is a two curve problem instance \( \mathcal{M} \), if \( h : \mathcal{M} \rightarrow \mathbb{C}^n \), we define its derivatives \( h^{(i)} \) in (C.5); for a kernel \( \Theta : \mathcal{M} \times \mathcal{M} \rightarrow \mathbb{R} \), we define its derivatives along the curve in Definition E.11.

For a Borel measure space \((X, \mu)\), a kernel \( K \) is a mapping \( K : X \times X \rightarrow \mathbb{R} \). We use \( K \) for its associated Fredholm integral operator. In other words, for measurable function \( f \) we have \( K(f)(x) = \int_{x' \in X} K(x, x') f(x') \, d\mu(x') \). When \( X \) is a Riemannian manifold, an omitted sub-/measure will always denote the Riemannian measure.

We use both lowercase and uppercase letters \( c, C \) for absolute constants whose value are independent of all parameters and \( c_{\tau}, C_{\tau} \) for numbers whose value only depend on some parameter \( \tau \). Throughout the text, \( c \) is used to represent numbers whose value should be small while \( C \) is used to represent numbers whose value only depend on some parameter \( \tau \). Throughout the text, \( \tau \) is used to denote application of the scalar function \( \tau \), with boundary condition \( x \rightarrow x^+ \) to each entry of the vector \( x \), with boundary condition \( \alpha^0_\Theta(x) = x \), and the network output on an input \( x \) is written \( f_\Theta(x) = W^{L+1}_\Theta \alpha^0_\Theta(x) \). We will also write \( \zeta_\Theta(x) = f_\Theta(x) - f_\star(x) \) to denote the fitting error. We use Gaussian initialization: if \( \ell \in \{1, 2, \ldots, L\} \), the weights are initialized as \( W^{\ell}_{ij} \sim_{\text{i.i.d.}} \mathcal{N}(0, 2/n) \), and the top level weights are initialized as \( W^{L+1}_i \sim_{\text{i.i.d.}} \mathcal{N}(0, 1) \) in order to preserve the expected feature norm.\(^3\) In the sequel, we will write \( \theta_0 \) to denote the collection of these initial random parameters, and therefore \( f_\theta_0 \) to denote the initial random network.

We will employ a convenient “empirical measure” notation to concisely represent finite-sample and population quantities in the analysis. Let \( \mu^N = \frac{1}{N} \sum_{i=1}^N \delta_{\{x_i\}} \) denote the empirical measure associated to our i.i.d. random sample from the population measure \( \mu \), where \( \delta_p \) denotes a Dirac measure at a point \( p \). We train on the square loss \( L^{\mu^N}(\Theta) = (1/2) \int_{\mathcal{M}} (\zeta_\Theta(x))^2 \, d\mu^N(x) \) (of course

\(^3\)This initialization style is common in practice (it might be referred to as “fan-out initialization” in that context), but less common in the theoretical literature on kernel regime training of deep neural networks, where a less-natural “NTK parameterization” is typically employed. A detailed discussion of these differences, and how to translate results for one parameterization into those for another, can be found (for example) in [6, §A.3].
one simply has $\mathcal{L}_{\mu}^N(\theta) = 1/(2N) \sum_{i=1}^{N} (\zeta_0(x_i))^2$, which we minimize using randomly-initialized “gradient descent” starting at $\theta_0$ with constant step size $\tau > 0$. We put gradient descent in quotations here because the loss $\mathcal{L}_{\mu}^N$ is only almost-everywhere differentiable, due to the nondifferentiability of the ReLU activation $\lceil \cdot \rceil_+$. Thus, in this sense our algorithm for minimization is ‘gradient-like’, in that it corresponds to a gradient descent iteration at almost all values of the parameters. Concretely, we define

$$\beta_{\theta}^{-1}(x) = (W_{L+1}^L \cdots W_{L-1}^L W_{L-2}^L \cdots W_{L+1}^2 W_{L+1}^1(x))$$
for $\ell = 0, 1, \ldots, L - 1$, where

$$I_{\ell}(x) = \{i \in [n] \mid \langle e_i, \alpha_{\theta}^\ell(x) \rangle > 0\}, \quad P_{I_{\ell}(x)} = \sum_{i \in I_{\ell}(x)} e_i e_i^\top$$
denotes the orthogonal projection onto the set of coordinates where the $\ell$-th activation at input $x$ is positive (above, $e_i$ denotes the $i$-th canonical basis vector, having its $j$-th entry equal to 1 if $j = i$ and 0 otherwise). Then we define ‘formal gradients’ of the network output with respect to the parameters (denoted by an operator $\nabla$) by

$$\nabla f_{\theta}(x) = \beta_{\theta}^{-1}(x) \alpha_{\theta}^{-1}(x)^*$$
for $\ell \in [L]$, and

$$\nabla L_{\mu}^N(\theta) = \int_M \nabla f_{\theta}(x) \zeta_0(x) \mu(x).$$

Thus, our gradient-like algorithm we study here is given by the sequence of parameters $\theta_{k+1} = \theta_k - \tau \nabla L_{\mu}^N(\theta_k)$, with $\theta_0$ given by the Gaussian initialization we describe above.

Our study of this gradient-like iteration is facilitated by using kernel regime techniques, which we will describe now. Formally, the gradient descent iteration implies the following “error dynamics” equation:

$$\zeta_{\theta_{k+1}}^N(x) = \zeta_{\theta_k}^N(x) - \tau \int_M \Theta_{\ell}^N(x, x') \zeta_{\theta_k}^N(x') \mu(x'),$$
where $\Theta_{\ell}^N(x, x') = \int_0^1 \langle \nabla f_{\theta_{k+1}}(x'), \nabla f_{\theta_{k+1}}(x) \rangle \zeta_{\theta_k}^N(x) \mu(x') \mu(x'')$. For a proof of this claim, see [6, Lemma B.8]. As we describe in Section 2.1, under suitable conditions on the network width, depth, and the number of samples, this error dynamics update is well-approximated by a “nominal dynamics” update equation defined by $\zeta_{k+1} = (1 - \tau \Theta_{\ell}^{\text{NTK}}) \zeta_k$ with boundary condition $\zeta_0 = \zeta_0$, where $\Theta_{\ell}^{\text{NTK}}(x, x') = \langle \nabla f_{\theta_0}(x), \nabla f_{\theta_0}(x') \rangle$ is the “neural tangent kernel”. The analysis of this nominal evolution leads us to the certificate problem that we have posed in Section 2.1, and which we resolve for the two curve problem in this work.

In the remainder of this section, we introduce several notations for quantities related to the certificate problem which we will refer to throughout these appendices. We let $\Theta$ denote the following approximation to the neural tangent kernel:

$$\Theta(x, x') = \sum_{\ell=0}^{L-1} \prod_{\ell' = \ell}^{L-1} \left( 1 - \frac{\varphi^{[\ell]}(\langle x, x' \rangle)}{\pi} \right),$$

where $\varphi^{[\ell]}$ denotes the $\ell$-fold composition of the angle evolution function $\varphi(t) = \cos^{-1}((1 - \frac{t}{\pi}) \cos t + \frac{\pi}{\pi} t)$. We let $\zeta$ denote the following piecewise constant approximation to $\zeta_0$:

$$\zeta(x) = -f_{\nu}(x) + \int_M f_{\theta_0}(x') \mu(x').$$
We also use the notation

$$\xi_{\ell}(t) = \sum_{\ell' = \ell}^{L-1} \left( 1 - \frac{\varphi^{[\ell]}(t)}{\pi} \right)$$
\[ \psi(t) = \frac{n}{2} \sum_{\ell=0}^{L-1} \xi_\ell(t) \]

for convenience. We find it convenient in our analysis to consider \( \psi \) and its “DC component”, i.e., its value at \( \pi \), separately. To this end, we write \( \psi^o = \psi - \psi(\pi) \). We also write the subtracted approximate NTK as \( \Theta^o(x, x') = \psi^o(\angle(x, x')) \). As a consequence, we have

\[ \psi^o(\angle(x, x')) = \Theta^o(x, x') = \Theta(x, x') - \psi(\pi). \] (C.4)

We use \( \Theta_\mu \) to represent the integral operator with

\[ \Theta_\mu[g](x) = \int_M \Theta(x, x') g(x') \, d\mu(x'), \]

and similarly for \( \Theta_\mu^o \). An omitted subscript/measure will denote the Riemannian measure on \( M \).

### C.2 Geometric Properties

We assume our data manifold \( M = M_+ \cup M_- \), where \( M_+ \) and \( M_- \) each is a smooth, regular, simple closed curve on the unit sphere \( S^{n_0-1} \). Because the curves are regular, it is without loss of generality to assume they are unit-speed and parameterized with respect to arc length, giving parameterizations as maps from \([0, \text{len}(M_p)] \) to \( S^{n_0-1} \), as we have defined them in Section 2.2 of the main body. Throughout the appendices, we will find it convenient to consider periodic extensions of these arc-length parameterizations, which are smooth and well-defined by the fact that our manifolds are smooth, closed curves: for \( \sigma \in \{\pm\} \), we use \( x_\sigma(s) : \mathbb{R} \rightarrow S^{n_0-1} \) to represent these parameterizations of the two manifolds.\(^4\) We require that the two curves are disjoint. Notice that as the two curves do not self intersect, we have \( x_\sigma(s) = x_{\sigma'}(s') \) if and only if \( \sigma = \sigma' \) and \( s' = s + k \text{len}(M_p) \) for some \( k \in \mathbb{Z} \). Precisely, our arguments will require our curves to have ‘five orders’ of smoothness, in other words \( x_\sigma(s) \) must be five times continuously differentiable for \( \sigma \in \{\pm\} \).

For a differentiable function \( h : M \rightarrow \mathbb{C}^p \) with \( p \in \mathbb{N} \), we define its derivative \( \frac{d}{ds}h \) as

\[
\frac{d}{ds}h(x) = \left[ \frac{d}{dt} \bigg|_{t=s} h(x_\sigma(t)) \right]_{x_\sigma(s)=x} \\
= \left[ \lim_{t \to 0} \frac{1}{t} \left( h(x_\sigma(s+t)) - h(x_\sigma(s)) \right) \right]_{x_\sigma(s)=x}.
\] (C.5)

We call attention to the “restriction” bar used in this notation: it should be read as “let \( s \) and \( \sigma \) be such that \( x_\sigma(s) = x \)” in the definition’s context. This leads to a valid definition in (C.5) because our curves are simple and disjoint, so for any choice \( s, s' \) with \( x_\sigma(s) = x_{\sigma'}(s') = x \), we have \( x_\sigma(s+t) = x_{\sigma'}(s'+t) \) for all \( t \). We will use this notation systematically throughout these appendices. We further denote its \( i \)-th order derivative by \( h^{(i)}(x) \). For \( i \in \mathbb{N} \), we use \( C^i(M) \) to represent the collection of real-valued functions \( h : M \rightarrow \mathbb{R} \) whose derivatives \( h^{(1)}, \ldots, h^{(i)} \) exist and are continuous.

In particular, consider the inclusion map \( \iota : M \rightarrow \mathbb{R}^{n_0} \), which is the identification \( \iota(x) = x \). Following the definition as above, we have

\[
\iota^{(i+1)}(x) = \left[ \lim_{t \to 0} \frac{1}{t} \left( \iota^{(i)}(x_\sigma(s+t)) - \iota^{(i)}(x_\sigma(s)) \right) \right]_{x_\sigma(s)=x}.
\] (C.6)

In the sequel, with abuse of notation we will use \( x^{(i)} \) to represent \( \iota^{(i)}(x) \). For example, we will write expressions such as \( \|x^{(2)}\|_2 \) to denote the quantity \( \sup_{x \in M} \|\iota^{(2)}(x)\|_2 \). This notation will

\(^4\)We clarify an abuse of notation we will commit with these parameterizations throughout the analysis, which stems from the fact that the curves are closed (i.e. topologically circles). That is, there is no preferred basepoint (i.e. the points \( x_\sigma(0) \)) for the arc length parameterizations (the curves are only defined up to translation); because our primary use for these parameterizations is in the analysis of extrinsic distances between points on the curves, the basepoint will be irrelevant.
We call the set $\mathcal{M}$, which are maps from $\mathbb{R}$ to $\mathbb{R}^n$ (say), rather than maps defined on $\mathcal{M}$. In this context, we also use $\dot{x}$ and $\ddot{x}$ to represent first and second derivatives $x^{(1)}$ and $x^{(2)}$ for brevity. We have $\|x\|_2 = \|\dot{x}\|_2 = 1$ from the fact that $\mathcal{M} \subset \mathbb{S}^{n-1}$ and that we have a unit-speed parameterization. This and associated facts are collected in Lemma E.3.

For any real or complex-valued function $h$, the integral operator over manifold can be written as

$$
\int_{x \in \mathcal{M}} h(x) d\mu(x) = \sum_{\sigma = \pm} \int_{s=0}^{\text{len}(\mathcal{M}_\sigma)} h(x_\sigma(s)) \rho(x_\sigma(s)) ds,
$$

$$
\int_{x \in \mathcal{M}} h(x) dx = \sum_{\sigma = \pm} \int_{s=0}^{\text{len}(\mathcal{M}_\sigma)} h(x_\sigma(s)) ds.
$$

We have defined key geometric properties in the main body, in Section 2.2. Our arguments will require slightly more technical definitions of these quantities, however. In the remainder of this section, we introduce the same definition of angle injectivity radius and $\mathcal{Q}$-number with a variable scale, which helps us in proofs in Appendix E.

First, we give a precise definition for the intrinsic distance $d_{\mathcal{M}}$ on the curves. To separate the notions of “close over the sphere” and “close over the manifold”, we use the extrinsic distance (angle) $\angle(x, x') = \cos^{-1} \langle x, x' \rangle$ to measures closeness between two points $x, x'$ over the sphere. The distance over the manifold is measured through the intrinsic distance $d_{\mathcal{M}}(x, x')$, which takes $\infty$ when $x$ and $x'$ reside on different components $\mathcal{M}_i$ and $\mathcal{M}_-$ and the length of the shortest curve on the manifold connecting the two points when they belong to the same component. More formally, we have

$$
d_{\mathcal{M}}(x, x') = \begin{cases} 
\inf \{|s - s'| : x_\sigma(s) = x, x_\sigma(s') = x', f_\sigma(x) = f_\sigma(x')\}, & \text{if and only if } s_1 = s_2 + k \text{ len}(\mathcal{M}_\sigma) \text{ for some } k \in \mathbb{Z},
\end{cases}
$$

where the infimum is taken over all valid $\sigma \in \{+, -\}$ and $(s, s') \in \mathbb{R}^2$. Notice that as the curves $\mathcal{M}_\sigma$ do not intersect themselves, one has $x_\sigma(s_1) = x_\sigma(s_2)$ if and only if $s_1 = s_2 + k \text{ len}(\mathcal{M}_\sigma)$ for some $k \in \mathbb{Z}$. Thus for any two points $x, x'$ that belong to the same component $\mathcal{M}_\sigma$, the above infimum is attained: there exist $s, s'$ such that $x_\sigma(s) = x, x_\sigma(s') = x'$, and $d_{\mathcal{M}}(x, x') = |s - s'|$.

**Angle Injectivity Radius** For $\varepsilon \in (0, 1)$ we define the angle injectivity radius of scale $\varepsilon$ as

$$
\Delta_\varepsilon = \min \left\{ \frac{\sqrt{\hat{k}}}{\kappa}, \inf_{x, x' \in \mathcal{M}} \left\{ \angle(x, x') \left| d_{\mathcal{M}}(x, x') \geq \frac{\sqrt{\varepsilon}}{\kappa} \right. \right\} \right\},
$$

which is the smallest extrinsic distance between two points whose intrinsic distance exceeds $\frac{\sqrt{\varepsilon}}{\kappa}$ with

$$
\hat{k} = \max \left\{ \kappa, \frac{2}{\pi} \right\}.
$$

Observe that for any scale $\varepsilon$, $\Delta_\varepsilon$ is smaller than inter manifold separation $\min_{x \in \mathcal{M}_+} d_{\mathcal{M}}(x, x')$.

**$\mathcal{Q}$-number** For $\varepsilon \in (0, 1), \delta \in (0, (1 - \varepsilon)]$, we define $\mathcal{Q}$-number of scale $\varepsilon, \delta$ as

$$
\mathcal{Q}_{\varepsilon, \delta}(\mathcal{M}) = \sup_{x \in \mathcal{M}} N_{\mathcal{M}} \left\{ \left\{ x' \left| d_{\mathcal{M}}(x, x') \geq \frac{\sqrt{\varepsilon}}{\kappa} \text{ and } \angle(x, x') \leq \frac{\delta \sqrt{\varepsilon}}{\kappa} \right. \right\}, \frac{1}{\sqrt{1 + \kappa^2}} \right\}.
$$

Here, $N_{\mathcal{M}}(T, \varepsilon)$ is the size of a minimal $\varepsilon$ covering of $T$ in the intrinsic distance on the manifold. We call the set $\left\{ x' \left| d_{\mathcal{M}}(x, x') \geq \frac{\sqrt{\varepsilon}}{\kappa}, \angle(x, x') \leq \frac{\delta \sqrt{\varepsilon}}{\kappa} \right. \right\}$ appearing in this definition the winding piece of scale $\varepsilon$ and $\delta$: it contains points that are far away in intrinsic distance but close in extrinsic distance. We will give it a formal definition in (E.6), where it will play a key role in our arguments.

In the sequel, we denote $\Delta, \mathcal{Q}(\mathcal{M})$ to be the angle injectivity radius and $\mathcal{Q}$-number with the specific instantiations $\varepsilon = \frac{1}{20}$ and $\delta = 1 - \varepsilon$. These are key geometric features used in Theorem 3.1 and Theorem 3.2.
C.3 Subspace of Smooth Functions and Kernel Derivatives

As the behavior of the kernel and its approximation is easier to understand when constrained in a low frequency subspace, we first introduce the notion of low-frequency subspace formed by the Fourier basis on the two curves.

Fourier Basis and Subspace of Smooth Functions

We define a Fourier basis of functions over the manifold as
\[ \phi_{\sigma,k}(x_{\sigma}(s)) = \begin{cases} \frac{1}{\sqrt{\text{len}(M_\sigma)}} \exp\left(\frac{i2\pi ks}{\text{len}(M_\sigma)}\right), & \sigma' = \sigma \\ 0, & \sigma' \neq \sigma \end{cases} \]  \hspace{1cm} (C.11)
for each \( k = 0, 1, \ldots \), and further define a subspace of low frequency functions
\[ S_{K_+, K_-} = \text{span}_C \{\phi_+, \phi_+, -1, \phi_+ + 1, \ldots, \phi_+, -K_+, \phi_+, K_+, \phi_-, 0, \ldots, \phi_-, K_-\} \]  \hspace{1cm} (C.12)
for \( K_+, K_- \geq 0 \). Using the fact that our curves are unit-speed, one can see that indeed (C.11) defines an orthonormal basis for \( L^2 \) functions on \( M \).

D Main Results

Theorem D.1 (Generalization). Let \( M \) be two disjoint smooth, regular, simple closed curves, satisfying \( \angle(x, x') \leq \pi/2 \) for all \( x, x' \in M \). For any \( 0 < \delta \leq 1/e \), choose \( L \) so that
\[ L \geq K \max \left\{ \left(\Delta (1 + \kappa^2)(1 + \kappa^{1/2})\right)^{20}, C_{\mu} \log^9 \left( \frac{1}{\delta} \right) \log^{24} \left( C_{\mu} n_0 \log (\frac{1}{\delta}) \right), c C^{\max\{\text{len}(M)\kappa, \text{log}(\kappa)\}}, P \right\} \]
\[ n = K' L'^{99} \log^{9} \left( 1/\delta \right) \log^{18} \left( L n_0 \right) \]
\[ N \geq L^{10}, \]
and fix \( \tau > 0 \) such that \( C'' \frac{\tau}{N^2} \leq \tau \leq \frac{c}{N}. \) Then with probability at least \( 1 - \delta \), the parameters obtained at iteration \( \|L_n\|_{44}/(n \tau) \) of gradient descent on the finite sample loss yield a classifier that separates the two manifolds.

The constants \( c, C, C', C'' \) are absolute, and the constant \( C_{\mu} \) is equal to \( \max\{\rho_{\min, \rho_{\max}}(1 + \rho_{\max})^{12} \} \). \( P \) is a polynomial \( \{M_1, M_2, M_3, \text{len}(M), \Delta^{-1}\} \) of degree at most \( 36 \), with degree at most \( 12 \) when viewed as a polynomial in \( M_1, M_2, M_3 \) and \( \text{len}(M) \), and of degree at most \( 24 \) as a polynomial in \( \Delta^{-1} \).

Proof. The proof is an application of Theorem G.1; we note that the conditions on \( n, L, \delta, N, \) and \( \tau \) imply all hypotheses of this theorem, except for the certificate condition. We will complete the proof by showing that the certificate condition is also satisfied, under the additional hypotheses on \( L \) and with a suitable choice of \( \tau \).

First, we navigate a difference in the formulation of the two curves’ regularity properties between our work and [6], from which Theorem G.1 is drawn. Theorem G.1 includes a condition \( L \geq C_K^\Delta C_\lambda \) for some absolute constant \( C \), where \( K^\Delta = \sup_{x, x' \in M} \|x - x'\|^2 \) is a bound on the extrinsic curvature (we will discuss \( \lambda \) momentarily). In our context, we have \( M_2 = K^\Delta \), and following Lemma E.3 (using that our curves are unit-speed spherical curves), we get that it suffices to require \( L \geq (1 + \kappa^2)C_\lambda \) instead. In turn, we can pass to \( \kappa \) since this constant is lower-bounded by a positive number and is larger than \( \kappa \), it suffices to require \( L \geq \kappa^2 C_\lambda \). As for \( C_\lambda \), this is a constant related to the angle injectivity radius \( \Delta \), and is defined by \( C_\lambda = K^2/\kappa^2 \), where these two constants satisfy
\[ \forall s \in (0, c_\lambda/\kappa^2], (x, x') \in M_+ \times M_+, * \in \{+, -\} : \angle(x, x') \leq s \Rightarrow d_M(x, x') \leq K_\lambda s. \]
We will relate this constant to constants in our formulation. Consider any \( x, x' \in M \). If \( \angle(x, x') \leq \frac{\Delta}{2} \) then from the definition of \( \Delta \) we have \( d_M(x, x') \leq \frac{\Delta}{\kappa} \) and hence by (E.31) we find \( d_M(x, x') \leq \angle(x, x') \). If on the other hand \( \angle(x, x') > \frac{\Delta}{2} \), then a trivial bound gives
\[ d_M(x, x') \leq \text{len}(M) = \frac{2\text{len}(M) \Delta}{\Delta} < \frac{2\text{len}(M) \Delta}{\Delta} \angle(x, x'). \]  \hspace{1cm} (D.1)
We can thus choose $c_\lambda = 1, K_\lambda = \max\left\{1, \frac{\text{len}(M)}{\Delta} \right\}$ to satisfy (D.1), giving $C_\lambda = \max\left\{1, \frac{\text{len}(M)}{\Delta^2} \right\}$. Thus the requirement $L > C_k^2 \Delta^2 C_\lambda$ of Theorem G.1 is automatically satisfied if $L \geq \max\{P, C\text{len}(M)^2\}$ for a suitable exponent $C$, where $P$ is the polynomial in the hypotheses of our result, and so our hypotheses imply this condition.

Next, we establish the certificate claim. The proof will follow closely the argument of [6, Proposition B.4]. Write $\Theta_{\text{NTK}}$ for the network’s neural tangent kernel, as defined in Appendix C.1, and $\Theta_{\mu,\nu}$ for the associated Fredholm integral operator on $L^2_{\nu}$. In addition, write $\zeta_0 = f_{\theta_0} - f_\star$ for the initial random network error. Because we have modified some exponents in the constant $C_\mu$, and added conditions on $L$, all hypotheses of Theorem D.2 are satisfied: invoking it, we have that there exists $g : \mathcal{M} \to \mathbb{R}$ satisfying

$$
\|g\|_{L^2_{\mu}} \leq C\frac{\|\zeta\|_{L^2_{\mu}}}{\rho_{\min}n}
$$

and

$$
\|\Theta_{\mu}[g] - \zeta\|_{L^2_{\mu}} \leq \frac{\|\zeta\|_{L^\infty}}{L}.
$$

By these bounds, the triangle inequality, the Minkowski inequality, and the fact that $\mu$ is a probability measure, we have

$$
\|\Theta_{\mu}^{\text{NTK}}[g] - \zeta_0\|_{L^2_{\mu}} \leq \|\Theta - \Theta_{\text{NTK}}\|_{L^\infty(\mathcal{M})}\|g\|_{L^2_{\mu}} + \|\Theta_{\mu}[g] - \zeta\|_{L^2_{\mu}} + \|\zeta - \zeta_0\|_{L^2_{\mu}}
$$

$$
\leq C\|\Theta - \Theta_{\text{NTK}}\|_{L^\infty(\mathcal{M})} \frac{\|\zeta\|_{L^\infty(\mathcal{M})} \sqrt{\text{len}(M)}}{n\rho_{\min}^2} + \frac{\|\zeta\|_{L^\infty}}{L} + \|\zeta - \zeta_0\|_{L^\infty(\mathcal{M})}.
$$

(D.2)

An application of Theorem G.2 gives that on an event of probability at least $1 - e^{-cd}$

$$
\|\Theta - \Theta_{\text{NTK}}\|_{L^\infty(\mathcal{M})} \leq Cn/L
$$

if $d \geq K \log(n\text{len}(M))$ and $n \geq K'd^4L^5$. In translating this result from [6], we use that in the context of the two curve problem, the covering constant $C_{\mathcal{M}}$ appearing in [6, Theorem B.2] is bounded by a constant multiple of $\text{len}(M)$ (this is how we obtain Theorem G.2 and some other results in Appendix G). An application of Lemma G.3 gives

$$
\mathbb{P}\left[\|\zeta_0 - \zeta\|_{L^\infty(\mathcal{M})} \leq \sqrt{2d} \leq \sqrt{\text{len}(\mathcal{M})} \leq \frac{\sqrt{d}}{L} \right] \geq 1 - e^{-cd}
$$

and

$$
\mathbb{P}\left[\|\zeta_0\|_{L^\infty(\mathcal{M})} \leq \sqrt{d} \right] \geq 1 - e^{-cd}
$$

as long as $n \geq K'd^4L^5$ and $d \geq K' \log(n\text{len}(M))$, where we use these conditions to simplify the residual that appears in Lemma G.3. In particular, combining the previous two bounds with the triangle inequality and a union bound and then rescaling $d$, which worsens the constant $c$ and the absolute constants in the preceding conditions, gives

$$
\mathbb{P}\left[\|\zeta\|_{L^\infty(\mathcal{M})} \leq \sqrt{d} \right] \geq 1 - 2e^{-cd}.
$$

Combining these bounds using a union bound and substituting into (D.2), we get that under the preceding conditions, on an event of probability at least $1 - 3e^{-cd}$ we have

$$
\|\Theta_{\mu}^{\text{NTK}}[g] - \zeta_0\|_{L^2_{\mu}} \leq C\frac{\sqrt{d}}{L} \left(2 + \frac{1}{\rho_{\min}}\right)
$$

$$
\leq C\frac{\sqrt{d}}{L} \max\{\rho_{\min}, \rho_{\min}^{-1}\}.
$$

(D.3)

where we worst-case the density constant in the second line, and in addition, on the same event, we have by the norm bound on the certificate $g$

$$
\|g\|_{L^2_{\mu}} \leq C\frac{\sqrt{d}}{n\rho_{\min}}.
$$

(D.4)
To conclude, we simplify the preceding conditions on $n$ and turn the parameter $d$ into a parameter $\delta > 0$ in order to obtain the form of the result necessary to apply Theorem G.1. We have in this one-dimensional setting
\[
\text{len}(\mathcal{M}) \leq \frac{\text{len}(\mathcal{M}_+)}{\mu(\mathcal{M}_+)} + \frac{\text{len}(\mathcal{M}_-)}{\mu(\mathcal{M}_-)} \leq \frac{2}{\rho_{\min}} \leq 2 \max\{\rho_{\min}, \rho_{\min}^{-1}\},
\]
where the second inequality here uses simply
\[
\mu(\mathcal{M}_+) = \int_{\mathcal{M}_+} \rho_+(x) \, dx \geq \text{len}(\mathcal{M}_+) \rho_{\min}
\]
(say). Because $n \geq 1$ and $n_0 \geq 3$ and $\max\{\rho_{\min}, \rho_{\min}^{-1}\} \geq 1$, it therefore suffices to instead enforce the condition on $d$ as $d \geq K \log(nn_0 C_{\mu})$, where $C_{\mu}$ is the constant defined in the lemma statement. But note from our hypotheses here that we have $n \geq L$ and $L \geq C_{\mu}$; so in particular it suffices to enforce $d \geq K \log(nn_0)$ for an adjusted absolute constant. Choosing $d \geq (1/e) \log(1/\delta)$, we obtain that the previous two bounds (D.3) and (D.4) hold on an event of probability at least $1 - 3\delta$. When $\delta \leq 1/e$, given that $n_0 \geq 3$ we have $nn_0 \geq e$ and $\max\{\log(1/\delta), \log(nn_0)\} \leq \log(1/\delta) \log(nn_0)$, so that it suffices to enforce the requirement $d \geq K \log(1/\delta) \log(nn_0)$ for a certain absolute constant $K > 0$. We can then substitute this lower bound on $d$ into the two certificate bounds above to obtain the form claimed in (G.1) in Theorem G.1 with the instantiation $q_{\text{cert}} = 1$, and this setting of $q_{\text{cert}}$ matches the choice of $C_{\mu}$ that we have enforced in our hypotheses here. For the hypothesis on $n$, we substitute this lower bound on $d$ into the condition on $n$ to obtain the sufficient condition $n \geq K'L^5 \log^4(1/\delta) \log^4(nn_0)$. Using a standard log-factor reduction (e.g. [6, Lemma B.15]) and possibly worsening absolute constants, we then get that it suffices to enforce $n \geq K'L^5 \log^4(1/\delta) \log^4(Ln_0 \log(1/\delta))$, which is redundant with the (much larger) condition on $n$ that we have enforced here. This completes the proof. \[\square\]

**Theorem D.2 (Certificates).** Let $\mathcal{M}$ be two disjoint smooth, regular, simple closed curves, satisfying $\mathcal{L}(\mathcal{M}, \mathcal{M}') \leq \pi/2$ for all $\mathcal{M}, \mathcal{M}' \subseteq \mathcal{M}$. There exist constants $C, C', C'', C'''$ and a polynomial $P = \text{poly}(M_3, M_4, M_5, \text{len}(\mathcal{M}), \Delta^{-1})$ of degree at most 36, with degree at most 12 in $(M_3, M_4, M_5, \text{len}(\mathcal{M}))$ and degree at most 24 in $\Delta^{-1}$, such that when
\[
L \geq \max\left\{\exp(C\text{len}(\mathcal{M})\tilde{\kappa}), \left(\frac{1}{\Delta \sqrt{1 + \kappa^2}}\right)^{C'''}\text{len}(\mathcal{M}), C'''K^{10}, P, \rho_{\max}^{12}\right\},
\]
then for $\zeta$ defined in (C.3), there exists a certificate $g : \mathcal{M} \to \mathbb{R}$ with
\[
\|g\|_{L^2} \leq \frac{C\|\zeta\|_{L^2}}{\rho_{\min} n \log L},
\]
such that
\[
\|\Theta_\mu[g] - \zeta\|_{L^2} \leq \|\zeta\|_{L^\infty} L^{-1}.
\]

**Proof.** This is a direct consequence of Lemma E.33. Notice that $\zeta$ in (C.3) is a real, piecewise constant function over the manifolds, and therefore has its higher order derivatives vanish. This makes it directly belong to $\Phi(\|\zeta\|_{L^2}, \frac{1}{\rho_{\min}})$ defined in Definition E.24 and satisfy the condition in Lemma E.33 with $K = 1$. \[\square\]

**E Proof for the Certificate Problem**

The goal of this section is to prove Lemma E.33, a generalized version of Theorem D.2. Instead of showing the certificate exists for the particular piecewise constant function $\zeta$ defined in (C.3), as claimed in Theorem D.2, Lemma E.33 claims that for any reasonably $\zeta$ with bounded higher order derivatives, there exists a small norm certificate $g$ such that $\Theta_\mu[g] \approx \zeta$. There are two main technical difficulties in establishing this result. First, $\Theta$ contains a very large constant term: $\Theta = \Theta^\circ + \psi(\pi) \Pi$. This renders the operator $\Theta$ somewhat ill-conditioned. Second, the eigenvalues of $\Theta^\circ$ are not bounded away from zero: because the kernel is sufficiently regular, it is possible to demonstrate high-frequency functions $h$ for which $\|\Theta^\circ h\|_{L^2} \ll \|h\|_{L^2}$.
Our proof handles these technical challenges sequentially: in Appendix E.1, we restrict attention to the DC subtracted kernel $\Theta^\circ$ and a subspace $S$ containing low-frequency functions, and show that the restriction $P_S \Theta^\circ P_S$ to $S$ is stably invertible over $S$. In Appendix E.2, we argue that the solution $g$ to $P_S \Theta^\circ [g] = \zeta$ is regularized enough that $\Theta^\circ [g] \approx \zeta$, i.e., the restriction to $S$ can be dropped. Finally, in Appendix E.3 we move from the DC subtracted kernel $\Theta^\circ$ without density to the full kernel $\Theta$. This move entails additional technical complexity; to maintain accuracy of approximation, we develop an iterative construction that successively applies the results of Appendix E.1–Appendix E.2 to whittle away approximation errors, yielding a complete proof of Lemma E.33.

E.1 Invertibility Over a Subspace of Smooth Functions

Proof Sketch and Organization. In this section, we solve a restricted version of the certificate problem for DC subtracted kernel $\Theta^\circ$, over a subspace $S$ of low-frequency functions defined in (C.12). Namely, for $\zeta \in S$, we demonstrate the existence of a small norm solution $g \in S$ to the equation

$$P_S \Theta^\circ [g] = \zeta. \tag{E.1}$$

This equation involves the integral operator $\Theta^\circ$, which acts via

$$\Theta^\circ [g] (x) = \int_{x' \in \mathcal{M}} \Theta^\circ (x, x') g(x') dx'. \tag{E.2}$$

We argue that this operator is invertible over $S$, by decomposing this integral into four pieces, which we call the Local, Near, Far, and Winding components. The formal definitions of these four components follow: for parameters $0 < \varepsilon < 1$, $r > 0$, and $\delta > 0$, we define

- **Local**: $L_r(x) = \{ x' \in \mathcal{M} \mid d_M(x, x') < r \}$, \hspace{1cm} (E.3)
- **Near**: $N_{r, \varepsilon}(x) = \{ x' \in \mathcal{M} \mid r \leq d_M(x, x') \leq \frac{\sqrt{\varepsilon}}{\hat{\kappa}} \}$, \hspace{1cm} (E.4)
- **Far**: $F_{\varepsilon, \delta}(x) = \{ x' \in \mathcal{M} \mid d_M(x, x') \geq \frac{\sqrt{\varepsilon}}{\hat{\kappa}}, \angle(x, x') > \frac{\delta \sqrt{\varepsilon}}{\hat{\kappa}} \}$, \hspace{1cm} (E.5)
- **Winding**: $W_{\varepsilon, \delta}(x) = \{ x' \in \mathcal{M} \mid d_M(x, x') \geq \frac{\sqrt{\varepsilon}}{\hat{\kappa}}, \angle(x, x') \leq \frac{\delta \sqrt{\varepsilon}}{\hat{\kappa}} \}$, \hspace{1cm} (E.6)

It is easy to verify that for any choice of these parameters and any $x \in \mathcal{M}$, these four pieces cover $\mathcal{M}$: i.e., $L_r(x) \cup N_{r, \varepsilon}(x) \cup F_{\varepsilon, \delta}(x) \cup W_{\varepsilon, \delta}(x) = \mathcal{M}$. Intuitively, the Local and Near pieces contain points that are close to $x$, in the intrinsic distance on $\mathcal{M}$. The Far component contains points that are far from $x$ in intrinsic distance, and far in the extrinsic distance (angle). The Winding component contains portions of $\mathcal{M}$ that are far in intrinsic distance, but close in extrinsic distance. Intuitively, this component captures parts of $\mathcal{M}$ that “loop back” into the vicinity of $x$.

Parameter choice. The specific parameters $r, \varepsilon, \delta$ will be chosen with an eye towards the properties of both $\mathcal{M}$ and $\Theta^\circ$. The parameter $\varepsilon \in (0, \frac{\pi}{4})$ is a scale parameter, which controls $r = r_\varepsilon$ such that

1. $r$ is large enough to enable the local component $L_r(x)$ to dominate the kernel’s behavior;
2. $r$ is not too large, so the kernel stays sharp and localized over the local component $L_r(x)$.

Specifically, we choose

$$a_\varepsilon = (1 - \varepsilon)^3 (1 - \varepsilon/12), \hspace{1cm} (E.7)$$

$$r_\varepsilon = 6\pi \tilde{L}^{-\frac{\varepsilon}{\pi + \varepsilon}}. \hspace{1cm} (E.8)$$

Notice that when $\varepsilon \searrow 0$, we have $r_\varepsilon \approx L^{-1/2}$. So with a smaller choice of $\varepsilon$ we may get a larger local component with the price of a larger constant dependence.

We further choose $\delta$ to ensure that the Near and Far components overlap. To see that this is possible, note that at the boundary of the Near component, $d_M(x, x') = \sqrt{\varepsilon}/\hat{\kappa}$; from Lemma E.4, we have

$$\angle(x, x') \geq d_M(x, x') - \hat{\kappa} \frac{\varepsilon}{\hat{\kappa}}, \hspace{1cm} (E.9)$$

so at this point $\angle(x, x') \geq (1 - \varepsilon)\sqrt{\varepsilon}/\hat{\kappa}$. Thus as long as $\delta < 1 - \varepsilon$, Near and Far overlap.
Kernel as main and residual. The kernel $\Theta^\circ(x, x')$ is a decreasing function of $\angle(x, x')$: $\Theta^\circ$ is largest over the Local component, smaller over the Near and Winding components, and smallest over the Far component. By choosing the scale parameter $r_\varepsilon$ as in (E.8), we define an operator $M_\varepsilon$ which captures the contribution of the Local component to the kernel:

$$
M_\varepsilon[f](x) = \int_{x' \in L_{r_\varepsilon}(x)} \psi^\circ(\angle(x, x')) f(x') \, dx'.
$$

(E.10)

Because $\angle(x, x')$ is small over $L_{r_\varepsilon}(x)$ when $r_\varepsilon$ is chosen to be small compared to inverse curvature $1/\kappa$, on this component, $dM_\varepsilon(x, x') \approx \angle(x, x')$ (which we formalize in Lemma E.4). We will use this property to argue that $M_\varepsilon$ can be approximated by a self-adjoint convolution operator, defined as

$$
\tilde{M}_\varepsilon[f](x) = \int_{s' = s-r_\varepsilon}^{s+r_\varepsilon} \psi^\circ(|s-s'|) f(x_\sigma(s')) \, ds' \bigg|_{x_\sigma(s) = x}.
$$

(E.11)

The restriction is valid because for any choice of $\sigma$ and $s$ such that $x_\sigma(s) = x$, the RHS has the same value. On the other hand, given that we require $0 < \varepsilon < \frac{2\pi}{3}$, (E.7) and (E.8) show that when $L$ is chosen larger than a certain absolute constant, we have $r_\varepsilon \leq \pi$, assuring $|s' - s|$ falls in the domain of $\psi^\circ$, which makes this operator well-defined. We will always assume such a choice has been made in the sequel, and in particular include it as a hypothesis in our results.

Notice that $\tilde{M}_\varepsilon$ is an invariant operator: it commutes with the natural translation action on $M$. As a result, it diagonalizes in the Fourier basis defined in (C.11) (i.e., each of these functions is an eigenfunction of $\tilde{M}_\varepsilon$). See Lemma E.6 and its proof for the precise formulation of these properties. This enables us to study its spectrum on the subspace of smooth functions defined in (C.12) at the specific scale $\varepsilon$, defined as

$$
S_\varepsilon = S_{K_{\varepsilon,+}, K_{\varepsilon,-}}
$$

(E.12)

with $K_{\varepsilon,\sigma} = \left[\varepsilon^{1/2} \text{len}(M_{\varepsilon,x})\kappa^{1/2}\right]$ for $\sigma \in \{+,-\}$. In this way, we will establish that $\tilde{M}_\varepsilon$ is stably invertible on $S_\varepsilon$.

In the remainder of the section, we show the diagonalizability and restricted invertibility of $\tilde{M}_\varepsilon$ in Lemma E.6, and control the $L^2$ to $L^2$ operator norm of all four components of $\Theta^\circ$ in Lemma E.7, Lemma E.8, Lemma E.9 and Lemma E.10. Then we show $\Theta^\circ$ is stably invertible using these results by a Neumann series construction (Lemma E.2) and finally prove the main theorem for this section in Theorem E.1.

**Theorem E.1.** For any $\varepsilon \in (0, \frac{\pi}{4})$, $\delta \in (0, 1 - \varepsilon]$, there exist an absolute constant $C$ and constants $C_{\varepsilon,\delta}, C', C''$ depending only on the subscripted parameters such that if

$$
L \geq \max \left\{ \exp\left(C_{\varepsilon,\delta} \text{len}(M)\kappa \right), \left(1 + \frac{1}{\Delta \sqrt{1 + \kappa^2}}\right)^{C_{\varepsilon,\delta} \text{len}(M)}, \varepsilon^{-1/2} 12 \pi \kappa^{a_\varepsilon + 1} \right\},
$$

where $a_\varepsilon, r_\varepsilon$ as in (E.7) and (E.8) and we set subspace $S_\varepsilon$ and the invariant operator $\tilde{M}_\varepsilon$ as in (E.12) and (E.11), we have $P_{S_\varepsilon} \tilde{M}_\varepsilon P_{S_\varepsilon}$ is invertible over $S_\varepsilon$, and

$$
\left\| \left( P_{S_\varepsilon} \tilde{M}_\varepsilon P_{S_\varepsilon} \right)^{-1} P_{S_\varepsilon} \left(\Theta^\circ - \tilde{M}_\varepsilon\right) P_{S_\varepsilon} \right\|_{L^2 \to L^2} \leq 1 - \varepsilon.
$$

Moreover, for any $\zeta \in S_\varepsilon$, the equation $P_{S_\varepsilon} \Theta^\circ \xi = \zeta$ has a unique solution $g_\varepsilon[\xi] \in S_\varepsilon$ given by the convergent Neumann series

$$
g_\varepsilon[\zeta] = \sum_{\ell=0}^{\infty} (-1)^\ell \left( \left( P_{S_\varepsilon} \tilde{M}_\varepsilon P_{S_\varepsilon} \right)^{-1} P_{S_\varepsilon} \left(\Theta^\circ - \tilde{M}_\varepsilon\right) P_{S_\varepsilon} \right)^\ell P_{S_\varepsilon} \tilde{M}_\varepsilon P_{S_\varepsilon}^{-1} \xi.
$$

(E.13)

Notice that although $\Theta$ and $\zeta$ are real objects, our subspace $S_\varepsilon$ contains complex-valued functions. In the remainder of Appendix E, we will work with complex objects for convenience, which means our constructed certificate candidates can be complex-valued. This will not affect our result because (intuitively) the fact that $\Theta$ and $\zeta$ are real makes the imaginary component of the certificate is redundant, and removing it with a projection onto the subspace of real-valued functions will give us the same norm and residual guarantees for the certificate problem. We make this claim rigorous and guarantee the existence of a real certificate in Lemma E.33, which is invoked in the proof of our main result on certificates, Theorem D.2.
which satisfies

$$\|g_\varepsilon[\zeta]\|_{L^2} \leq \frac{C\|\zeta\|_{L^2}}{\varepsilon n \log L}. \quad (E.14)$$

**Proof.** We construct $g \in S_\varepsilon$ satisfying $P_{S_\varepsilon}\Theta^\circ[g] = \zeta$ by equivalently writing

$$P_{S_\varepsilon}\Theta^\circ[g] = \left( P_{S_\varepsilon} \tilde{M}_\varepsilon P_\varepsilon + P_{S_\varepsilon} \left( \Theta^\circ - \tilde{M}_\varepsilon \right) P_{S_\varepsilon} \right) [g].$$

Under our hypotheses, Lemma E.6 implies the invertibility of $P_{S_\varepsilon} \tilde{M}_\varepsilon P_{S_\varepsilon}$ with

$$\lambda_{\min} \left( P_{S_\varepsilon} \tilde{M}_\varepsilon P_{S_\varepsilon} \right) \geq \frac{1}{1 - \varepsilon} \left\| \Theta^\circ - \tilde{M}_\varepsilon \right\|_{L^2 \to L^2}, \quad (E.15)$$

where $\lambda_{\min} \left( P_{S_\varepsilon} \tilde{M}_\varepsilon P_{S_\varepsilon} \right)$ is the minimum eigenvalue of the self-adjoint operator $P_{S_\varepsilon} \tilde{M}_\varepsilon P_{S_\varepsilon} : S_\varepsilon \to S_\varepsilon$ as shown in Lemma E.6. In particular, $P_{S_\varepsilon} \tilde{M}_\varepsilon P_{S_\varepsilon}$ is invertible, and the system we seek to solve can be written equivalently as

$$\left( P_{S_\varepsilon} \tilde{M}_\varepsilon P_{S_\varepsilon} \right)^{-1} \zeta = \left( \text{Id}_{S_\varepsilon} + \left( P_{S_\varepsilon} \tilde{M}_\varepsilon P_{S_\varepsilon} \right)^{-1} P_{S_\varepsilon} \left( \Theta^\circ - \tilde{M}_\varepsilon \right) P_{S_\varepsilon} \right) [g],$$

where the LHS of the last system is in $S_\varepsilon$. Next, we argue that the operator that remains on the RHS of the last equation is invertible. Noting that

$$\left\| \left( P_{S_\varepsilon} \tilde{M}_\varepsilon P_{S_\varepsilon} \right)^{-1} P_{S_\varepsilon} \left( \Theta^\circ - \tilde{M}_\varepsilon \right) P_{S_\varepsilon} \right\|_{L^2 \to L^2} \leq \lambda_{\min} \left( P_{S_\varepsilon} \tilde{M}_\varepsilon P_{S_\varepsilon} \right)^{-1} \left\| \Theta^\circ - \tilde{M}_\varepsilon \right\|_{L^2 \to L^2} \leq 1 - \varepsilon \quad (E.16)$$

using both Lemma E.6 and (E.15), we have by the Neumann series that

$$\left( \text{Id}_{S_\varepsilon} + \left( P_{S_\varepsilon} \tilde{M}_\varepsilon P_{S_\varepsilon} \right)^{-1} P_{S_\varepsilon} \left( \Theta^\circ - \tilde{M}_\varepsilon \right) P_{S_\varepsilon} \right)^{-1} = \sum_{i=0}^{\infty} (-1)^i \left( P_{S_\varepsilon} \tilde{M}_\varepsilon P_{S_\varepsilon} \right)^{-1} P_{S_\varepsilon} \left( \Theta^\circ - \tilde{M}_\varepsilon \right) P_{S_\varepsilon}^i.$$

Thus we know $g_\varepsilon[\zeta]$ in (13) serves as the solution to the equation $P_{S_\varepsilon}\Theta^\circ[g] = \zeta$.

Furthermore, from Lemma E.6 when $L \geq C_\varepsilon$, we have

$$\left\| \left( P_{S_\varepsilon} \tilde{M}_\varepsilon P_{S_\varepsilon} \right)^{-1} \zeta \right\|_{L^2} \leq \lambda_{\min} \left( P_{S_\varepsilon} \tilde{M}_\varepsilon P_{S_\varepsilon} \right)^{-1} \|\zeta\|_{L^2} \leq \frac{1}{\varepsilon n \log L} \|\zeta\|_{L^2}.$$

Combining this bound with (E.16) and the triangle inequality in the series representation (E.13), we obtain the claimed norm bound in (E.14):

$$\|g_\varepsilon[\zeta]\|_{L^2} \leq \sum_{i=0}^{\infty} (1 - \varepsilon)^i \left\| \left( P_{S_\varepsilon} \tilde{M}_\varepsilon P_{S_\varepsilon} \right)^{-1} \zeta \right\|_{L^2} \leq \frac{C\|\zeta\|_{L^2}}{\varepsilon n \log L},$$

$\blacksquare$
Lemma E.2. Let $\varepsilon \in (0, \frac{3}{4})$, $\delta \in (0, 1 - \varepsilon]$, and let $a_\varepsilon$, $\hat{M}_\varepsilon$ and $S_\varepsilon$ be as in (E.7), (E.11) and (E.12). There are constants $C_\varepsilon, C_{\varepsilon, \delta}, C''_{\varepsilon, \delta}$ depending only on the subscripted parameters such that if

$$L \geq \max \left\{ \exp \left( C'_{\varepsilon, \delta} \text{len}(\mathcal{M}) \frac{1}{K} \right), \left( 1 + \frac{1}{\Delta \sqrt{1 + K^2}} \right)^{C''_{\varepsilon, \delta}} (\varepsilon^{-1/2} 12\pi \hat{k})^{\frac{a_\varepsilon + 1}{a_\varepsilon}} C_\varepsilon \right\},$$

we have $P_{S_\varepsilon} \hat{M}_\varepsilon P_{S_\varepsilon}$ is invertible over $S_\varepsilon$ with

$$\lambda_{\min} \left( P_{S_\varepsilon} \hat{M}_\varepsilon P_{S_\varepsilon} \right) \geq \frac{1}{1-\varepsilon} \left\| \Theta^0 - \hat{M}_\varepsilon \right\|_{L^2 \rightarrow L^2}
$$

where $\lambda_{\min} \left( P_{S_\varepsilon} \hat{M}_\varepsilon P_{S_\varepsilon} \right)$ is defined in Lemma E.6.

Proof. From triangle inequality for the $L^2 \rightarrow L^2$ operator norm, we have

$$\left\| \Theta^0 - M_\varepsilon \right\|_{L^2 \rightarrow L^2} \leq \left\| \Theta^0 - M_\varepsilon \right\|_{L^2 \rightarrow L^2} + \left\| M_\varepsilon - \hat{M}_\varepsilon \right\|_{L^2 \rightarrow L^2}.
$$

To bound the first term, we define

$$M_\varepsilon(x, x') = \mathbb{1}_{d_{\mathcal{M}}(x, x') \leq r_{\varepsilon}} \psi^\varepsilon(\angle(x, x')).$$

Then it is a bounded symmetric kernel $\mathcal{M} \times \mathcal{M} \rightarrow \mathbb{R}$, and following (E.10), $M_\varepsilon$ is its associated Fredholm integral operator. We can thus apply Lemma E.5 and get

$$\left\| \Theta^0 - M_\varepsilon \right\|_{L^2 \rightarrow L^2} \leq \sup_{x \in \mathcal{M}} \int_{x' \in \mathcal{M}} |\Theta^0(x, x') - M_\varepsilon(x, x')| dx' = \sup_{x \in \mathcal{M}} \int_{x' \in \mathcal{M} \setminus L_{r_{\varepsilon}}(x)} |\Theta^0(x, x')| dx' + \int_{x' \in L_{r_{\varepsilon}}(x)} |\Theta^0(x, x')| dx'.
$$

Because the Near, Far and Winding pieces cover $\mathcal{M} \setminus L_{r_{\varepsilon}}(x)$, we have

$$\int_{x' \in \mathcal{M} \setminus L_{r_{\varepsilon}}(x)} |\Theta^0(x, x')| dx' \leq \int_{x' \in N_{r_{\varepsilon}}(x)} |\Theta^0(x, x')| dx' + \int_{x' \in W_{r_{\varepsilon}}(x)} |\Theta^0(x, x')| dx' + \int_{x' \in F_{r_{\varepsilon}}(x)} |\Theta^0(x, x')| dx'.
$$

From Lemma E.7, Lemma E.8, Lemma E.9 and Lemma E.10, we know that there exist constants $C_2, C_3, C_4$ and for any $\varepsilon'' \leq 1$ exists numbers $C_{\varepsilon'', C''_{\varepsilon, \delta}}$ such that when $L \geq C_{\varepsilon''}$ and $L \geq (\varepsilon^{-1/2} 12\pi \hat{k})^{\frac{a_\varepsilon + 1}{a_\varepsilon}}$, we have

$$\left\| \Theta^0 - \hat{M}_\varepsilon \right\|_{L^2 \rightarrow L^2} \leq \sup_{x \in \mathcal{M}} \int_{x' \in N_{r_{\varepsilon}}(x)} |\Theta^0(x, x')| dx' + \sup_{x \in \mathcal{M}} \int_{x' \in W_{r_{\varepsilon}}(x)} |\Theta^0(x, x')| dx' + \sup_{x \in \mathcal{M}} \int_{x' \in F_{r_{\varepsilon}}(x)} |\Theta^0(x, x')| dx' + \left\| \hat{M}_\varepsilon - M_\varepsilon \right\|_{L^2 \rightarrow L^2} \leq \frac{3\pi n}{4(1-\varepsilon)} \left( 1 + \varepsilon'' \right) \log \left( \frac{\sqrt{\varepsilon}}{K r_{\varepsilon}} \right) + C_2, n
$$

$$+ C_2 \text{len}(\mathcal{M}) \frac{\hat{k}}{\sqrt{\varepsilon}}
$$

$$+ C_3 \sup_{x, \delta} \text{len}(\mathcal{M}) n \log \left( 1 + \frac{1}{\Delta \sqrt{1 + K^2}} \right)
$$

$$+ C_4 (1 - \varepsilon)^{-2} \hat{k}^2 r_{\varepsilon}^2.
$$

(16.17)

Meanwhile, from Lemma E.6 there exists constant $C_\varepsilon, C_1$ such that when $L \geq C_\varepsilon$

$$(1 - \varepsilon) \lambda_{\min} \left( P_{S_\varepsilon} \hat{M}_\varepsilon P_{S_\varepsilon} \right) \geq (1 - \varepsilon)^2 \frac{3\pi n}{4} \log \left( 1 + \frac{L - 2}{3\pi r_{\varepsilon}} \right) - C_1 (1 - \varepsilon) n r_{\varepsilon} \log^2 L.
$$

(16.18)
We will treat all named constants appearing in the previous two equations as fixed for the remainder of the proof. We argue that the first term in this expression is large enough to dominate each of the terms in (E.17) and the residual term in (E.18).

Set $\epsilon' = \frac{\epsilon}{2\pi}$. We will choose $\epsilon'' = \frac{\epsilon'}{2\pi} < 1$, so that both $\epsilon'$ and $\epsilon''$ depend only on $\epsilon$. Then, since $r_\epsilon = 6\pi L^{-\frac{\epsilon}{2\pi+\epsilon}}$, when $L > 4$, we have
\[
\frac{L-2}{3\pi} r_\epsilon = 2(L-2)L^{-\frac{\epsilon}{2\pi+\epsilon}} > L^{\frac{1}{2\pi+\epsilon}}. \tag{E.19}
\]
Since moreover $a_\epsilon = (1 - \epsilon)^3(1 - \frac{\epsilon'}{2\pi}) = (1 - \epsilon)^3(1 - 2\epsilon')$, we have $a_\epsilon \epsilon'' = \epsilon'(1 - \epsilon)^3$, and therefore $(1 + \epsilon'')a_\epsilon = (1 - \epsilon')(1 - \epsilon)^3$. Thus
\[
(1 - \epsilon')(1 - \epsilon)^2 \frac{3\pi n}{4} \log \left(1 + \frac{L - 2}{3\pi} r_\epsilon \right) = \frac{3\pi n}{4(1 - \epsilon)} (1 + \epsilon'') a_\epsilon \log \left(1 + \frac{L - 2}{3\pi} r_\epsilon \right) \geq \frac{3\pi n}{4(1 - \epsilon)} (1 + \epsilon'') \log \left(L^{\frac{n}{2\pi+\epsilon}} \right) \geq \frac{3\pi n}{4(1 - \epsilon)} (1 + \epsilon'') \log \left(\frac{\sqrt{\epsilon}}{\kappa r_\epsilon} \right), \tag{E.20}
\]
where in the last bound we use $\sqrt{\epsilon} \kappa^{-1} \leq 6\pi$, given that $\epsilon < 1$ and $\kappa \leq \pi/2$. The RHS at the end of this chain of inequalities is the first term of the RHS of the last bound in (E.17). Since the LHS has a leading coefficient of $(1 - \epsilon')$, we can conclude provided we can split the remaining $\epsilon'$ across the remaining terms.

Next, we will cover the negative term in (E.18) and the second and fifth terms in (E.17). Using (E.19), we have
\[
\frac{\epsilon'}{3} (1 - \epsilon)^2 \frac{3\pi n}{4} \log \left(1 + \frac{L - 2}{3\pi} r_\epsilon \right) \geq \frac{\epsilon'}{3} (1 - \epsilon)^2 \frac{3\pi n}{4} \frac{1}{a_\epsilon + 1} \log(L). \tag{E.21}
\]
There exists a constant $C_\epsilon$ such that when $L \geq C_\epsilon$, we have for the RHS
\[
\frac{\epsilon'}{3} (1 - \epsilon)^2 \frac{3\pi n}{4} \frac{1}{a_\epsilon + 1} \log(L) \geq (C_1 + C_4 + C_{\epsilon''}) n.
\]
In particular, we can take
\[
C_\epsilon \geq \exp \left( \frac{C_1 + C_4 + C_{\epsilon''}}{\epsilon' (1 - \epsilon)^2 \frac{3\pi n}{4} \frac{1}{a_\epsilon + 1}} \right).
\]
Next, there exists another constant $C_\epsilon > 0$ such that when $L \geq C_\epsilon$, we have $r_\epsilon \log^2 L \leq 1$, whence by the previous bound
\[
\frac{\epsilon'}{3} (1 - \epsilon)^2 \frac{3\pi n}{4} \frac{1}{a_\epsilon + 1} \log(L) \geq (C_1 r_\epsilon \log^2 L + C_4 + C_{\epsilon''}) n.
\]
Finally, notice that when $L \geq (\epsilon^{-1/2} 12\pi \kappa) \frac{n + 1}{n}$, we have
\[
r_\epsilon = 6\pi L^{-\frac{\epsilon}{2\pi+\epsilon}} \leq \frac{\sqrt{\epsilon}}{2\kappa},
\]
so $r_\epsilon \kappa \leq \sqrt{\epsilon}/2$, and since $\epsilon \in (0, 3/4)$, we have
\[
(1 - \epsilon) C_1 n r_\epsilon \log^2 L + C_4 (1 - \epsilon)^{-2} \kappa^2 n r_\epsilon^2 + C_{\epsilon''} n \leq (C_1 r_\epsilon \log^2 L + 3C_4 + C_{\epsilon''}) n,
\]
where we used that $\epsilon \mapsto (1 - \epsilon)^{-2}$ is increasing. Combining our previous bounds, this gives
\[
\frac{\epsilon'}{3} (1 - \epsilon)^2 \frac{3\pi n}{4} \log \left(1 + \frac{L - 2}{3\pi} r_\epsilon \right) \geq (1 - \epsilon) C_1 n r_\epsilon \log^2 L + C_4 (1 - \epsilon)^{-2} \kappa^2 n r_\epsilon^2 + C_{\epsilon''} n, \tag{E.22}
\]
as desired.

For the remaining two terms, define
\[
C_{\epsilon, \delta} = \frac{(a_\epsilon + 1) C_2}{(1 - \epsilon)^2 \frac{3\pi n}{4} \delta}, \quad C_{\epsilon''} = \frac{(a_\epsilon + 1) C_3}{(1 - \epsilon)^2 \frac{3\pi n}{4}}.
\]
We will use the estimate (E.21) as our base. Then when
\[ L \geq \max \left\{ \exp \left( C_{\varepsilon, \delta} \text{len}(\mathcal{M}) \hat{\kappa} \right), \left( 1 + \frac{1}{\Delta \varepsilon \sqrt{1 + \kappa^2}} \right) C''_{\varepsilon, \delta} \text{len}(\mathcal{M}) \right\}, \]
we have
\[ \frac{\varepsilon'}{3} (1 - \varepsilon)^2 \frac{3 \pi n}{4} \log \left( 1 + \frac{L - 2}{3 \pi} r_\varepsilon \right) \geq C_2 \text{len}(\mathcal{M}) \frac{n}{\delta \sqrt{\varepsilon}}, \quad (E.23) \]
\[ \frac{\varepsilon'}{3} (1 - \varepsilon)^2 \frac{3 \pi n}{4} \log \left( 1 + \frac{L - 2}{3 \pi} r_\varepsilon \right) \geq C_3 \text{len}(\mathcal{M}) \frac{n \log \left( 1 + \frac{1}{\Delta \varepsilon \sqrt{1 + \kappa^2}} \right)}{\delta \sqrt{\varepsilon}}, \quad (E.24) \]
Combining (E.20), (E.22), (E.23), (E.24) completes the proof. ■

Lemma E.3. For any \( x \in \mathcal{M} \), we have
\[ \langle x, \dot{x} \rangle = \langle \dot{x}, \ddot{x} \rangle = 0, \quad (E.25) \]
\[ \langle x, x^{(3)} \rangle = -1, \quad (E.26) \]
\[ \langle x, x^{(4)} \rangle = -\langle \dot{x}, x^{(3)} \rangle = ||\dot{x}||_2^2, \quad (E.27) \]
\[ \langle \ddot{x}, x^{(3)} \rangle = -\frac{1}{3} \langle \dot{x}, x^{(4)} \rangle, \quad (E.28) \]
\[ \| P_{x^\perp} \ddot{x} \|_2^2 = ||\dot{x}||_2^2 - 1, \quad (E.29) \]
where we use above the notation introduced near (C.6).

Proof. As our curve is defined over sphere and has unit speed, we have
\[ ||x||_2^2 = ||\dot{x}||_2^2 = 1. \]
Taking derivatives on both sides, we get
\[ \langle x, \dot{x} \rangle = \langle \dot{x}, \ddot{x} \rangle = 0. \]
Continuing to take higher derivatives, we get the following relationships:
\[ ||\dot{x}||_2^2 + \langle x, \ddot{x} \rangle = 0, \]
\[ 3 \langle \dot{x}, \ddot{x} \rangle + \langle x, x^{(3)} \rangle = 0, \]
\[ 3 ||\ddot{x}||_2^2 + 4 \langle \dot{x}, x^{(3)} \rangle + \langle x, x^{(4)} \rangle = 0, \]
\[ ||\ddot{x}||_2^2 + \langle \ddot{x}, x^{(3)} \rangle = 0, \]
\[ 3 \langle \ddot{x}, x^{(3)} \rangle + \langle \dot{x}, x^{(4)} \rangle = 0. \]
which gives us by plugging in the previous constraints
\[ \langle x, \dot{x} \rangle = -1, \]
\[ \langle x, x^{(3)} \rangle = 0, \]
\[ \langle \dot{x}, x^{(3)} \rangle = -||\dot{x}||_2^2, \]
\[ \langle x, x^{(4)} \rangle = ||\dot{x}||_2^2, \]
\[ \langle \dot{x}, x^{(4)} \rangle = -\frac{1}{3} \langle \dot{x}, x^{(4)} \rangle. \]
As a consequence, the intrinsic curvature \(\|P_{x^+} \dot{x}\|_2\) and extrinsic curvature \(\|x\|_2\) are related by
\[
\|P_{x^+} \dot{x}\|_2^2 = \| (I - xx^*) \dot{x} \|_2^2 = (x, \dot{x})^2 + (\dot{x}, \dot{x}) - 2 (x, \dot{x})^2 = \|\dot{x}\|_2^2 - 1.
\]
Thus we know
\[
M_2 = \sup_{x \in M} \|\dot{x}\|_2
= \sup_{x \in M} \sqrt{1 + \|P_{x^+} \dot{x}\|_2^2}
= \sqrt{1 + \sup_{x \in M} \left\{ \|P_{x^+} \dot{x}\|_2 \right\}^2}
\leq \sqrt{1 + \kappa^2}
\leq \sqrt{\left( \frac{\pi \kappa}{2} \right)^2 + \kappa^2}
< 2 \hat{\kappa}.
\]
Furthermore, the above shows that \(M_2 \geq 1\), so we have
\[
M_2 \leq M_2^2 = \sup_{x \in M} \|\dot{x}\|_2^2
= \sup_{x \in M} \left\langle x, x^{(4)} \right\rangle
\leq M_4,
\]
using one of our previously-derived relationships in the second line and Cauchy-Schwarz in the third. Finally, for any point \(x = x_\sigma(s)\), as \(x_\sigma(s + \text{len}(\mathcal{M}_\sigma)) = x_\sigma(s)\), we have
\[
0 = x_\sigma(s + \text{len}(\mathcal{M}_\sigma)) - x_\sigma(s) = \int_{s'=s}^{s+\text{len}(\mathcal{M}_\sigma)} \dot{x}_\sigma(s')ds'
= \text{len}(\mathcal{M}_\sigma) \dot{x}_\sigma(s) + \int_{s'=s}^{s+\text{len}(\mathcal{M}_\sigma)} \int_{s''=s'}^{s''} \ddot{x}_\sigma(s'')ds''ds'
\]
which leads to
\[
\text{len}(\mathcal{M}_\sigma) = \left\| \int_{s'=s}^{s+\text{len}(\mathcal{M}_\sigma)} \int_{s''=s'}^{s''} \ddot{x}_\sigma(s'')ds''ds' \right\|_2
= \left\| \int_{s'=s}^{s+\text{len}(\mathcal{M}_\sigma)} \int_{s''=s'}^{s''} \ddot{x}_\sigma(s'')ds''ds' \right\|_2
\leq \int_{s'=s}^{s+\text{len}(\mathcal{M}_\sigma)} \int_{s''=s'}^{s''} M_2 ds''ds'
= \frac{\text{len}(\mathcal{M}_\sigma)^2}{2} M_2 < \text{len}(\mathcal{M}_\sigma)^2 \hat{\kappa},
\]
completing the proof, where the first line uses the unit-speed property, the second uses the previous relation, the third uses Jensen’s inequality (given that \(\| \cdot \|_2\) is convex and 1-homogeneous), and the last line comes from (E.28).

**Lemma E.4.** Let \(\hat{\kappa} = \max \left\{ \kappa, \frac{2}{\pi} \right\}\). For \(\sigma \in \{\pm\}\) and \(|s' - s| \leq \frac{1}{\hat{\kappa}}\), we have
\[
|s - s'| - \hat{\kappa}|s - s'|^3 \leq \angle(x_\sigma(s), x_\sigma(s')) \leq |s - s'|.
\]
(E.30)
As a consequence, for \(|s - s'| \leq \frac{\sqrt{\pi}}{\kappa}\),
\[
(1 - \varepsilon)|s - s'| \leq \angle(x_\sigma(s), x_\sigma(s')) \leq |s - s'|.
\]
(E.31)
In particular, for any two points \(x, x' \in \mathcal{M}_\sigma\), choosing \(s, s'\) such that \(x_\sigma(s) = x, x_\sigma(s') = x'\), and \(|s - s'| = d_M(x, x')\), we have when \(d_M(x, x') \leq \frac{\sqrt{\pi}}{\kappa}\)
\[
(1 - \varepsilon)d_M(x, x') \leq \angle(x, x') \leq d_M(x, x').
\]
Proof. We prove (E.30) first.

The upper bound is direct from the fact that $\mathcal{M}$ is a pair of paths in the sphere and $\angle(x, x')$ is the length of a path in the sphere of minimum distance between points $x, x'$, and then using the fact that the distance $|s' - s| \geq d_{\mathcal{M}}(x_{s}(s), x_{s}(s'))$ from (C.7).

The lower bound requires some additional estimates. We fix $s, s'$ satisfying our assumptions; as both $|s - s'|$ and $\angle(x_{s}(s), x_{s}(s'))$ are symmetric functions of $(s, s')$, it suffices to assume that $s' \geq s$. Define $t = s' - s$, then by assumption we have $0 \leq t \leq \frac{1}{2} \leq \frac{s}{2}$. As $\cos^4$ is strictly decreasing on $[-1, 1]$, we only need to show that

$$\langle x_{s}(s), x_{s}(s + t) \rangle \leq \cos(t - \kappa^2 t^3). \quad (E.32)$$

Using the second order Taylor expansion at $s$, we have

$$x_{s}(s + t) = x_{s}(s) + \dot{x}_{s}(s) + \int_{a=s}^{s+t} \ddot{x}_{s}(b) db \, da$$

and so

$$\langle x_{s}(s), x_{s}(s + t) \rangle = \langle x_{s}(s), x_{s}(s) + \dot{x}_{s}(s) + \int_{a=s}^{s+t} \ddot{x}_{s}(b) db \, da \rangle$$

$$= \|x_{s}(s)\|^2 + \langle x_{s}(s), \dot{x}_{s}(s) \rangle + \langle x_{s}(s), \int_{a=s}^{s+t} \ddot{x}_{s}(b) db \, da \rangle$$

$$= 1 + \int_{a=s}^{s+t} \int_{b=s}^{a} \langle x_{s}(s), \ddot{x}_{s}(b) \rangle db \, da \quad (E.33)$$

where we use properties established in Lemma E.3, in particular (E.25) in the last line. Take second order Taylor expansion at $b$ for $x_{s}(s)$, we have similarly

$$x_{s}(s) = x_{s}(b) + \dot{x}_{s}(b) + \int_{c=s}^{b} \ddot{x}_{s}(d) dd \, dc.$$ 

From (E.25) and (E.26), we have $\langle x_{s}(b), \ddot{x}_{s}(b) \rangle = -1$ and $\langle \dot{x}_{s}(b), \ddot{x}_{s}(b) \rangle = 0$. Thus uniformly for $b \in [s, s + t]$

$$\langle x_{s}(s), \ddot{x}_{s}(b) \rangle = -1 + \int_{c=s}^{b} \int_{d=c}^{b} \ddot{x}_{s}(d) dd \, dc \, db$$

$$= -1 + \int_{c=s}^{b} \int_{d=c}^{b} \langle \ddot{x}_{s}(d), \ddot{x}_{s}(b) \rangle dd \, dc$$

$$\leq -1 + \int_{c=s}^{b} \int_{d=c}^{b} \|\ddot{x}_{s}(d)\|_{2} \|\ddot{x}_{s}(b)\|_{2} dd \, dc$$

$$\leq -1 + \int_{c=s}^{b} \int_{d=c}^{b} M_{s}^2 dd \, dc$$

$$\leq -1 + \frac{M_{s}^2}{2} (b - s)^2,$$

where in the third line we use Cauchy-Schwarz. Plugging this last bound into (E.33), it follows

$$\langle x_{s}(s), x_{s}(s + t) \rangle \leq 1 + \int_{a=s}^{s+t} \int_{b=s}^{a} (-1 + (M_{s}^2/2)(b - s)^2) db \, da$$

$$= 1 - \frac{t^2}{2} + \frac{M_{s}^2}{2} \int_{a=s}^{s+t} \int_{b=s}^{a} (b - s)^2 db \, da$$

$$= 1 - \frac{t^2}{2} + \frac{M_{s}^2}{4!} t^4$$

$$= 1 - \frac{t^2}{2} + \frac{1 + \kappa^2 t^2}{4!} t^4,$$  \quad (E.34)
with an application of Lemma E.3 in the final equality. To conclude, we derive a suitable estimate for \( \cos(t - \hat{\kappa}^2 t^3) \). Because \( 0 \leq t \leq \hat{\kappa}^{-1} \), we have that \( t^{-1}(t - \hat{\kappa}^2 t^3) \in [0, 1] \), and because \( t \leq \hat{\kappa}^{-1} \leq \pi/2 \), we can apply concavity of \( \cos \) on \([0, \pi/2] \) to obtain

\[
\cos(t - \hat{\kappa}^2 t^3) \geq \frac{t - \hat{\kappa}^2 t^3}{t} \cos(t) + \left(1 - \frac{t - \hat{\kappa}^2 t^3}{t}\right) \cos(0).
\]

Next, the estimate \( \cos(x) \geq 1 - \frac{x^2}{2} + \frac{x^4}{4!} - \frac{x^6}{6!} \) for all \( x \), a consequence of Taylor expansion, gives

\[
(1 - \hat{\kappa}^2 t^2) \cos(t) + \hat{\kappa}^2 t^2 \geq (1 - \hat{\kappa}^2 t^2) \left(1 - \frac{t^2}{2} + \frac{t^4}{4!} - \frac{t^6}{6!}\right) + \hat{\kappa}^2 t^2 = 1 - \frac{t^2}{2} + \frac{\hat{\kappa}^2 t^4}{4!} - \frac{\hat{\kappa}^2 t^6}{6!} + \hat{\kappa}^2 t^8
\]

after distributing. Because \( \hat{\kappa} \geq \kappa \), we can split terms and write

\[
t^4/4! + \hat{\kappa}^2 t^4/2 \geq \frac{1 + \kappa^2}{4!} t^4 + \hat{\kappa}^2 t^4/4,
\]

and then grouping terms in the preceding estimates gives

\[
\cos(t - \hat{\kappa}^2 t^3) \geq 1 - \frac{t^2}{2} + \frac{1 + \kappa^2}{4!} t^4 + \hat{\kappa}^2 t^4 \left(1 - \frac{t^2}{4!} + \frac{t^4}{6!} - \frac{t^2}{6\hat{\kappa}^2}\right).
\]

By way of (E.34) and (E.32), we will therefore be done if we can show that

\[
\frac{1}{4} - \left(\frac{1}{4!} + \frac{1}{6\hat{\kappa}^2}\right) t^2 + \frac{t^4}{6!} \geq 0.
\]

This is not hard to obtain: for example, we can prove the weaker but sufficient bound

\[
1 - \frac{1}{3!} \left(1 + \frac{1}{30\kappa^2}\right) t^2 \geq 0
\]

by noticing that because \( t \leq \hat{\kappa}^{-1} \), it suffices to show

\[
\frac{1}{\hat{\kappa}^2} \left(1 + \frac{1}{30\kappa^2}\right) \leq 6,
\]

and because the LHS of the previous line is an increasing function of \( \hat{\kappa}^{-1} \) and moreover \( \hat{\kappa}^{-1} \leq \pi/2 \), this bound follows by verifying that indeed \((\pi/2)^2((1 + (1/30)(\pi/2)^2)\leq 6 \). Because \( s, s'\) were arbitrary we have thus proved (E.30).

For the remaining claims, (E.31) follows naturally from the fact that when \( |s - s'| \leq \sqrt{\pi} \), we have \( |s - s'| - \hat{\kappa}^2|s - s'|^3 \geq (1 - \varepsilon)|s - s'| \). The final claim is a restatement of (E.31) under the additional stated hypotheses.

Invertibility of \( \hat{M} \) over \( S \).

Lemma E.5 (Young’s inequality for Fredholm operators). Let \( K : \mathcal{M} \times \mathcal{M} \to \mathbb{R} \) satisfy \( K(x, x') = K(x', x) \) for all \( (x, x') \in \mathcal{M} \times \mathcal{M} \) and \( \sup_{(x, x') \in \mathcal{M} \times \mathcal{M}} |K(x, x')| < +\infty \), and let \( K \) denote its Fredholm integral operator (defined as \( g \mapsto K[g] = \int_{\mathcal{M}} K(\cdot, x') g(x') \, dx' \)). For any \( 1 \leq p \leq +\infty \), we have

\[
\|K\|_{L^p \to L^p} \leq \sup_{x \in \mathcal{M}} \int_{x' \in \mathcal{M}} |K(x, x')| \, dx'.
\]

Proof. The proof uses the M. Riesz convexity theorem for interpolation of operators [1, §V, Theorem 1.3], which we need here in the form of a special case: it states that for all \( 1 \leq p \leq +\infty \), one has

\[
\|K\|_{L^p \to L^p} \leq \|K\|_{L^\infty \to L^\infty}^{1/p} \|K\|_{L^1 \to L^1}^{1-1/p}. \tag{E.35}
\]

To proceed, we will bound the two operator norm terms on the RHS. We have

\[
\|K\|_{L^1 \to L^1} = \sup_{\|g\|_{L^1} = 1} \int_{x \in \mathcal{M}} \int_{x' \in \mathcal{M}} K(x, x') g(x') \, dx' \, dx
\]
As a consequence, which shows that each Fourier basis function is an eigenfunction of $\hat{P}_c$. These two bounds are equal; plugging them into (E.35) thus proves the claim.

In the first inequality above uses the triangle inequality for the integral. In the third line, we rearrange the order of integration using Fubini’s theorem, given that $g$ is integrable and $K$ is bounded on $M \times M$. In the fourth line, we use $L^1 - L^\infty$ control of the integrand (i.e., Hölder’s inequality), and in the final line we use that $\|g\|_{L^1} = 1$ along with symmetry of $K$ and nonnegativity of the integrand to re-index and remove the outer absolute value. On the other hand, $L^1 - L^\infty$ control and the triangle inequality give immediately

$$\|K\|_{L^\infty \to L^\infty} = \sup_{x \in M, \|g\|_{L^\infty} = 1} \left| \int_{x' \in M} K(x, x') g(x') \, dx' \right|$$

$$\leq \sup_{x \in M, x' \in M} \int_{x' \in M} |K(x, x')| \, dx'.$$

These two bounds are equal; plugging them into (E.35) thus proves the claim. 

**Lemma E.6.** Let $\varepsilon \in (0, \frac{3}{4})$, $r_\varepsilon$, $S_\varepsilon$ and $\hat{M}_\varepsilon$ be as defined in (E.8), (C.12) and (E.11). Then $\hat{M}_\varepsilon$ diagonalizes in the Fourier orthonormal basis (C.11). Write $\lambda_{\min} \left( P_{S_\varepsilon} \hat{M}_\varepsilon P_{S_\varepsilon} \right)$ for the minimum eigenvalue of the operator $P_{S_\varepsilon} \hat{M}_\varepsilon P_{S_\varepsilon} : S_\varepsilon \to S_\varepsilon$. Then there exist constants $c$ and a constant $C_\varepsilon$ such that when $L \geq C_\varepsilon$, we have

$$\lambda_{\min} \left( P_{S_\varepsilon} \hat{M}_\varepsilon P_{S_\varepsilon} \right) \geq (1 - \varepsilon) \frac{3\pi n}{4} \log \left( 1 + \frac{L - 2}{3\pi} r_\varepsilon \right) - Cn r_\varepsilon \log^2 L,$$

$$\geq cn \log L.$$

As a consequence, $P_{S_\varepsilon} \hat{M}_\varepsilon P_{S_\varepsilon}$ is invertible over $S_\varepsilon$, and

$$\left\| \left( P_{S_\varepsilon} \hat{M}_\varepsilon P_{S_\varepsilon} \right)^{-1} \right\|_{L^2 \to L^2} = \lambda_{\min}^{-1} \left( P_{S_\varepsilon} \hat{M}_\varepsilon P_{S_\varepsilon} \right).$$

**Proof.** Choose $L \geq 1$ to guarantee that $\hat{M}_\varepsilon$ is well-defined. We use $\psi^\circ$ to denote the DC subtracted skeleton, as defined in (C.4), and $(\phi_{\sigma,k})_{\sigma,k}$ the (intrinsic) Fourier basis on $M$, as defined in (C.11). For any Fourier basis function $\psi^\circ_{\sigma,k}$, we have

$$\hat{M}_\varepsilon [\phi_{\sigma,k}(x_\sigma)] = \int_{s' = -r_\varepsilon}^{s' = r_\varepsilon} \psi^\circ (|s - s'|) \phi_{\sigma,k}(x_\sigma (s')) \, ds'$$

$$= \int_{s' = -r_\varepsilon}^{s' = r_\varepsilon} \psi^\circ (|s'|) \exp \left( \frac{i2\pi ks'}{\text{len}(M_\sigma)} \right) \, ds' \phi_{\sigma,k}(x_\sigma (s))$$

$$= \phi_{\sigma,k}(x_\sigma (s)) \int_{s' = -r_\varepsilon}^{s' = r_\varepsilon} \psi^\circ (|s'|) \cos \left( \frac{2\pi ks'}{\text{len}(M_\sigma)} \right) \, ds',$$

which shows that each Fourier basis function is an eigenfunction of $\hat{M}_\varepsilon$; because these functions form an orthonormal basis for $L^2(M)$ (by classical results from Fourier analysis on the circle), $\hat{M}_\varepsilon$ diagonalizes in this basis. Moreover, because $S_\varepsilon$ is the span of Fourier basis functions, $P_{S_\varepsilon}$ also diagonalizes in this basis, and hence so does $P_{S_\varepsilon} \hat{M}_\varepsilon P_{S_\varepsilon}$. Because $\hat{M}_\varepsilon$ is self-adjoint and $P_{S_\varepsilon}$ is an orthogonal projection, $P_{S_\varepsilon} \hat{M}_\varepsilon P_{S_\varepsilon}$ is self-adjoint; and because $\text{dim}(S_\varepsilon) < +\infty$, the operator $P_{S_\varepsilon} \hat{M}_\varepsilon P_{S_\varepsilon}$ has finite rank, and therefore has a well-defined minimum eigenvalue, which we denote
as in the statement of the lemma. As \( K_{\varepsilon, \sigma} = \left\lfloor \frac{\varepsilon^{1/2} \text{len}(M_\sigma)}{2\pi r_\varepsilon} \right\rfloor \), we have for any \(|k_\sigma| \leq K_{\varepsilon, \sigma}\) and any \(|s'| \leq r_\varepsilon\),

\[
1 \geq \cos \left( \frac{2\pi k_\sigma s'}{\text{len}(M_\sigma)} \right) \geq 1 - \left( \frac{2\pi k_\sigma s'}{\text{len}(M_\sigma)} \right)^2 \geq 1 - \varepsilon.
\]

Then for \( \sigma \in \{+, -\} \) and \(|k| \leq K_{\varepsilon, \pm}\),

\[
\tilde{M}_\varepsilon[\phi_{\sigma,k}(x_\sigma(s))] = \phi_{\sigma,k}(x_\sigma(s)) \int_{s' = -r_\varepsilon}^{r_\varepsilon} \psi^\sigma(|s'|) \cos \left( \frac{2\pi ks'}{\text{len}(M_\sigma)} \right) ds',
\]

and so

\[
\lambda_{\min}(P_{S_\varepsilon} \tilde{M}_\varepsilon P_{S_\varepsilon}) \geq 2(1 - \varepsilon) \int_0^{r_\varepsilon} \psi^\sigma(s) ds.
\]

From Lemma E.7, we have if \( L \geq 1 \)

\[
2 \int_{s=0}^{r_\varepsilon} \psi^\sigma(s) ds \geq \frac{3\pi n}{4} \log \left( 1 + \frac{L - 2}{3\pi} r_\varepsilon \right) - Cnr_\varepsilon \log^2 L.
\]

In particular, as \( r_\varepsilon = 6\pi L^{-\frac{n}{n+2\pi}} \), there exists a constant \( C_\varepsilon \) such that when \( L \geq C'_\varepsilon \), we have

\[
Cnr_\varepsilon \log^2 L \leq \frac{\varepsilon}{4} \frac{3\pi n}{4} \log \left( 1 + \frac{L - 2}{3\pi} r_\varepsilon \right),
\]

and thus

\[
\lambda_{\min}(P_{S_\varepsilon} \tilde{M}_\varepsilon P_{S_\varepsilon}) \geq (1 - \varepsilon) \left( 1 - \frac{\varepsilon}{4} \right) \frac{3\pi n}{4} \log \left( 1 + \frac{L - 2}{3\pi} r_\varepsilon \right)
\]

\[
\geq \left( 1 - \frac{5\varepsilon}{4} \right) \frac{3\pi n}{4} \log \left( L^{1 - \frac{n}{n+2\pi}} \right)
\]

\[
= \left( 1 - \frac{5\varepsilon}{4} \right) \frac{3\pi n}{4} \frac{1}{a_\varepsilon + 1} \log L
\]

\[
\geq \left( 1 - \frac{5\varepsilon}{4} \right) \frac{3\pi n}{4} \frac{1}{2} \log L
\]

\[
\geq cn \log L
\]

\[> 0,
\]

where we used \( L \geq 1 \) in the second inequality, and \( \varepsilon < 3/4 \) and \( a_\varepsilon \leq 1 \) in the third inequality. So \( P_{S_\varepsilon} \tilde{M}_\varepsilon P_{S_\varepsilon} \) is invertible over \( S_\varepsilon \), with

\[
\left( P_{S_\varepsilon} \tilde{M}_\varepsilon P_{S_\varepsilon} \right)^{-1} [h] = \sum_{\sigma = \pm} \sum_{k = 0}^{K_{\varepsilon, \sigma}} \left( \int_{s = -r_\varepsilon}^{r_\varepsilon} \psi^\sigma(|s|) \cos \left( \frac{2\pi ks}{\text{len}(M_\sigma)} \right) ds \right)^{-1} \phi_{\sigma,k} \phi_{\sigma,k}^* h. \quad (E.37)
\]

The final claim is a consequence of the fact that \( P_{S_\varepsilon} \tilde{M}_\varepsilon P_{S_\varepsilon} \) is self-adjoint and finite-rank.

\[\blacksquare\]

**Lemma E.7.** Let \( \varepsilon \in (0, \frac{3}{4}) \), \( a_\varepsilon, r_\varepsilon, M_\varepsilon \) and \( \tilde{M}_\varepsilon \) be as defined in (E.7), (E.8), (E.10) and (E.11). There exist constants \( C, C' \), such that when \( L \geq C \) and \( L \geq (\varepsilon^{-1/2} 12\pi \tilde{K})^\frac{n+1}{n+2\pi} \), we have

\[
\left\| M_\varepsilon - \tilde{M}_\varepsilon \right\|_{L^2 \to L^2} \leq (1 - \varepsilon)^{-2} C' \tilde{K}^2 nr_\varepsilon^2.
\]

**Proof.** We choose \( L \geq 1 \) to guarantee that \( \tilde{M}_\varepsilon \) is well-defined for all \( 0 < \varepsilon < 3/4 \). We would like to use Lemma E.5 to bound \( \left\| M_\varepsilon - \tilde{M}_\varepsilon \right\|_{L^2 \to L^2} \), and thus we define two (suggestively-named) bounded symmetric kernels \( M_\varepsilon \times M_\varepsilon \to \mathbb{R} \):

\[
M_\varepsilon(x, x') = \mathbb{1}_{d_M(x, x') < r_\varepsilon} \psi^\sigma(\angle(x, x'))
\]
and
\[
\hat{M}_e(x, x') = \mathbb{1}_{d_M(x, x') < r_e} \psi^o(d_M(x, x')).
\]
From (E.10), \( M_e \) is indeed \( M_e \)'s associated Fredholm integral operator. To show that under our constraints for \( L \), \( \hat{M}_e \) is also \( \hat{M}_e \)'s associated integral operator, we first notice that following (C.7), for any \( x, x' \in M, d_M(x, x') < r_e \) if and only if there exist \( \sigma, s \) and \( s' \) such that \( x = x_\sigma(s), x' = x_\sigma(s') \) and \( |s' - s| < r_e \). This means for any fixed \( x \), if we let \( \sigma, s \) be chosen such that \( x = x_\sigma(s) \), then \( L_{r_e}(x) = \{ x_\sigma(s') ||s' - s| < r_e \} \). Furthermore, as \( L \geq (\varepsilon^{-1/2} 12 \pi \kappa)^{\frac{1}{12}} \), by (E.29) in Lemma E.3 we have \( r_e \leq \frac{\sqrt{2}}{2 \kappa} < \min \{ \text{len}(M_+), \text{len}(M_-) \}/2 \). Under this condition, we can unambiguously express the intrinsic distance \( d_M \) in terms of arc length at the local scale: for any \( x' \in L_{r_e}(x) \), there is a unique \( s' \) such that \( |s' - s| \leq r_e \). To see this, note that for any other parameter choice that attains the infimum in (C.7) \( s'' = s' + k \text{len}(M_\sigma) \) with integer \( k \neq 0 \), the triangle inequality implies \( |s'' - s| \geq |r_e - k \text{len}(M_\sigma)| \), and one has \( |r_e - k \text{len}(M_\sigma)| > r_e \) for every \( k \neq 0 \) if \( 0 < r_e < \text{len}(M_\sigma)/2 \). Then for \( x' \in L_{r_e}(x) \) and any \( s' \in [s - r_e, s + r_e] \) such that \( x_\sigma(s') = x' \), we have \( d_M(x, x') = |s - s'| \). Combining all these points, \( \hat{M}_e \)'s associated Fredholm integral operator \( H \) can be written as:
\[
H[f](x_\sigma(s)) = \int_{d_M(x_\sigma(s), x') < r_e} \psi^o(d_M(x_\sigma(s), x')) f(x') dx'
\]
\[
= \int_{x' \in \{ x_\sigma(s)| |s'-s| < r_e \}} \psi^o(d_M(x_\sigma(s), x')) f(x') dx'
\]
\[
= \int_{s - r_e}^{s + r_e} \psi^o(d_M(x_\sigma(s), x_\sigma(s'))) f(x_\sigma(s')) ds'
\]
\[
= \int_{s' = s - r_e}^{s' = s + r_e} \psi^o(|s - s'| f(x_\sigma(s'))) ds'
\]
\[
= \hat{M}_e[f](x_\sigma(s)),
\]
which means \( \hat{M}_e \) is indeed \( \hat{M}_e \)'s associated integral kernel.

We can now apply Lemma E.5 and and get
\[
\| M_e - \hat{M}_e \|_{L^2 \rightarrow L^2} \leq \sup_{x \in M} \int_{x' \in M} |M_e(x, x') - \hat{M}_e(x, x')| dx'
\]
\[
= \sup_{x \in M} \int_{x' \in L_{r_e}(x)} |M_e(x, x') - \hat{M}_e(x, x')| dx'
\]
\[
= \sup_{s, \sigma} \int_{s' = s - r_e}^{s' = s + r_e} \left| \psi^o \left( \angle(x_\sigma(s), x_\sigma(s')) \right) - \psi^o \left( |s - s'| \right) \right| ds'. \quad (E.38)
\]
Here, we recall that \( r_e < \pi/4 \) (because \( \kappa \leq \pi/2 \)), so there is no issue with these evaluations and the domain of \( \psi^o \) being \([0, \pi]\). Note that from (E.30), when \( |s - s'| \leq r_e \leq \frac{\sqrt{2}}{2 \kappa} \), we have
\[
\angle(x_\sigma(s), x_\sigma(s')) \geq |s - s'| - \kappa^2 |s - s'|^3 \\
\geq (1 - \varepsilon) |s - s'|. \quad (E.39)
\]
As \( \psi^o \) is nonnegative, strictly decreasing and convex by Lemma G.5, we know both \( \psi^o \) and \( |\psi^o| \) are decreasing. Also, by the upper bound in Lemma E.4, we have that \( \psi^o(\angle(x_\sigma(s), x_\sigma(s'))) - \psi^o(|s - s'|) \geq 0 \), so we can essentially ignore the absolute value in the integrand in (E.38). We can then calculate
\[
\int_{s' = s - r_e}^{s' = s + r_e} \psi^o \left( \angle(x_\sigma(s), x_\sigma(s')) \right) - \psi^o \left( |s - s'| \right) ds'
\]
\[
\leq \int_{s' = s - r_e}^{s' = s + r_e} \psi^o \left( |s - s'| - \kappa^2 |s - s'|^3 \right) - \psi^o \left( |s - s'| \right) ds'
\]
\[
= \int_{t = -r_e}^{r_e} \psi^o \left( |t| - \kappa^2 |t|^3 \right) - \psi^o \left( |t| \right) dt
\]

24
where in the last line we apply a linear change of variables. We also note that in the above integrals above, the first line comes from (E.39) and the fact that $\psi$ is strictly decreasing and that in (E.39). The last line uses symmetry and a linear transformation. Note that from (E.39) we always have $|t| - \hat{\kappa}^2|t^3|$ nonnegative when $|t| \leq r_x$ and thus all above formulas are well defined. From Lemma E.10, we know that there exists $C, C'$ such that when $L \geq C$, we have

$$\int_0^{(1-\varepsilon)r_x} t^3|\psi(t)|dt \leq C'n(1-\varepsilon)^2r_x^2,$$

and plugging all bounds back to (E.38) we get

$$\|M_\varepsilon - \hat{M}_\varepsilon\|_{L^2 \to L^2} \leq (1-\varepsilon)^{-2}C'n^2\varepsilon^2$$

as claimed.

Lemma E.8. Let $\varepsilon \in (0, \frac{3}{4})$, $r_\varepsilon$ and $N_{r_\varepsilon,\varepsilon}$ as defined in (E.8) and (E.4). For any $0 < \varepsilon'' \leq 1$, there exist numbers $C'_\varepsilon, C''_{\varepsilon,\varepsilon}$ such that when $L \geq C''_{\varepsilon,\varepsilon}$, we have

$$\sup_{x \in \mathcal{M}} \int_{x' \in N_{r_\varepsilon,\varepsilon}(x)} \psi^0(\angle(x, x')) ds' \leq \frac{3\pi n}{4(1-\varepsilon)}(1+\varepsilon'') \log \left(\frac{\sqrt{\varepsilon}}{\hat{\kappa}r_x}\right) + C''_{\varepsilon,\varepsilon}n.$$

Proof. For $x \in \mathcal{M}$, assume the parameters are chosen such that the corresponding near piece is nonempty, for otherwise the claim is immediate. Recalling (E.4), for any $x' \in N_{r_\varepsilon,\varepsilon}(x)$, we have $d_{\mathcal{M}}(x, x') \leq \sqrt{\varepsilon}/\hat{\kappa}$. From Lemma E.4, this implies $\angle(x, x') \geq (1-\varepsilon)d_{\mathcal{M}}(x, x')$. Let $\sigma, s$ be such that $x_{\sigma}(s) = x$. Notice by the discussion following the definition of the intrinsic distance in (C.7) that the near component $N_{r_\varepsilon,\varepsilon}(x)$ is contained in the set $\{x_{\sigma}(s') \mid |s' - s| \leq r_x, \sqrt{\varepsilon}/\hat{\kappa}\}$. And from Lemma G.5, $\psi^0$ is strictly decreasing, thus we have

$$\int_{x' \in N_{r_\varepsilon,\varepsilon}(x)} \psi^0(\angle(x, x')) ds' \leq \int_{s' = s + r_x}^{s + \sqrt{\varepsilon}} \psi^0(\angle(x_{\sigma}(s), x_{\sigma}(s'))) ds' \leq \int_{s' = s - \sqrt{\varepsilon}}^{s - r_x} \psi^0(\angle(x_{\sigma}(s), x_{\sigma}(s'))) ds' \leq \int_{s' = s - \sqrt{\varepsilon}}^{s + \sqrt{\varepsilon}} \psi^0((1-\varepsilon)|s' - s|) ds' \leq \int_{s' = s - \sqrt{\varepsilon}}^{s + \sqrt{\varepsilon}} \psi^0((1-\varepsilon)|s' - s|) ds' \leq \int_{s' = s - \sqrt{\varepsilon}}^{s + \sqrt{\varepsilon}} \psi^0((1-\varepsilon)|s' - s|) ds' = 2 \int_{t = r_x}^{(1-\varepsilon)|s|} \psi^0((1-\varepsilon)t) dt = 2 \int_{t = (1-\varepsilon)r_x}^{(1-\varepsilon)|s|} \psi^0(t) dt,$$

where in the last line we apply a linear change of variables. We also note that in the above integrals $|s' - s| \leq \hat{\kappa}^{-1} \leq \pi/2$, so there are no issues above with the domain of $\psi^0$ being $[0, \pi]$. From
Lemma E.9, for any \(0 < \varepsilon'' \leq 1\), there exist numbers \(C_{\varepsilon''}, C'_{\varepsilon''}\) such that if \(L \geq C_{\varepsilon''}\), then \(r_x\) satisfies the condition in (F.12) and we have
\[
\sup_{x \in \mathcal{M}} \int_{x' \in N_{r_x}(x)} \psi^\circ \left( \angle(x, x') \right) \, ds' \leq \frac{2}{1 - \varepsilon} \int_{t = (1 - \varepsilon)r_x} \frac{1}{n} \psi^\circ (|t|) \, dt \\
\leq \frac{2}{1 - \varepsilon} (1 + \varepsilon'') \frac{3\pi n}{8} \log \left( \frac{1 + (L - 3)(1 - \varepsilon)r_x/(3\pi)}{1 + (L - 3)(1 - \varepsilon)r_x/(3\pi)} \right) \\
+ C'_{\varepsilon''} n \\
\leq \frac{3\pi n}{4(1 - \varepsilon)} (1 + \varepsilon'') \log \left( \frac{\sqrt{\varepsilon}}{\kappa \varepsilon} \right) + C'_{\varepsilon''} n.
\]

\[\blacksquare\]

Lemma E.9. Let \(\varepsilon \in (0, \frac{1}{4})\), \(\delta \in (0, 1 - \varepsilon]\). Let \(W_{\varepsilon, \delta}\) as in (E.6). There exist constants \(C, C'\) such that when \(L \geq C\), for any \(x \in \mathcal{M}\),
\[
\int_{x' \in W_{\varepsilon, \delta}(x)} \psi^\circ \left( \angle(x, x') \right) \, ds' \leq \bar{C} \varepsilon_{\varepsilon, \delta}(\mathcal{M}) C' \log \left( 1 + \frac{\sqrt{\varepsilon}}{\Delta x} \right) .
\]

**Proof.** To bound the integral, we rely on the observation that for each ‘curve segment’ inside the winding component, the angle \(\angle(x, x')\) cannot stay small for the whole segment, and thus we can avoid worst case control for the angle as we have employed for the far component in Lemma E.10.\(^6\)

We will begin by constructing a specific finite cover of curve segments for the winding component, then we will bound the integral over each curve segment by providing a lower bound for the angle function.

As \(\mathcal{M}\) is compact with bounded length, from the definition in (C.10) we know \(\bar{C} \varepsilon_{\varepsilon, \delta}(\mathcal{M})\) is a finite number for any choice of \(\varepsilon, \delta\). From the definition of the winding component (E.6), for any point \(x \in \mathcal{M}\), we can cover \(W_{\varepsilon, \delta}(x)\) by at most \(\bar{C} \varepsilon_{\varepsilon, \delta}(\mathcal{M})\) closed balls in the intrinsic distance on the manifold with radii no larger than \(1/\sqrt{1 + \kappa^2}\). Topologically, each ball in the intrinsic distance of radii \(r\) is a curve segment of length \(2r\); thus, \(W_{\varepsilon, \delta}(x)\) can be covered by at most \(2\bar{C} \varepsilon_{\varepsilon, \delta}(\mathcal{M})\) curve segments, each with length no larger than \(1/\sqrt{1 + \kappa^2}\). Formally, this implies that for each \(x \in \mathcal{M}\), there exists a number \(N(x) \leq 2\bar{C} \varepsilon_{\varepsilon, \delta}(\mathcal{M})\) and for each \(i \in \{1, \cdots, N(x)\}\), there exist a sign \(\sigma_i(x) \in \{\pm\}\) and a nonempty interval \(I_i(x) = [s_{1,i}(x), s_{2,i}(x)]\) with length no greater than \(\frac{1}{\sqrt{1 + \kappa^2}}\) and strictly less than \(\text{len}(\mathcal{M}_{\sigma_i(x)})\) such that
\[
W_{\varepsilon, \delta}(x) \subseteq \bigcup_{i=1}^{N(x)} X_i(x)
\]
where \(X_i(x) = \{x_{\sigma_i(x)}(s) \mid s \in I_i(x)\} \subseteq \mathcal{M}\) with \(X_i(x) \cap W_{\varepsilon, \delta}(x) \neq \emptyset\). For the purpose of minimum coverage, we can further assume without loss of generality that for each \(x\) and each \(i\), the boundary points \(x_{\sigma_i(x)}(s_{1,i}(x))\) and \(x_{\sigma_i(x)}(s_{2,i}(x))\) belong to \(W_{\varepsilon, \delta}(x)\): we can always set \(p_{1,i}(x) = \inf\{s \mid s \in [s_{1,i}(x), s_{2,i}(x)], x_{\sigma_i(x)}(s) \in W_{\varepsilon, \delta}(x)\}\) and \(p_{2,i}(x) = \sup\{s \mid s \in [s_{1,i}(x), s_{2,i}(x)], x_{\sigma_i(x)}(s) \in W_{\varepsilon, \delta}(x)\}\), then the curve segment associated with \(\sigma_i(x)\) and interval \([p_{1,i}(x), p_{2,i}(x)]\) still covers \(X_i(x) \cap W_{\varepsilon, \delta}(x)\). As \(W_{\varepsilon, \delta}(x)\) is closed, we have the boundary points \(x_{\sigma_i(x)}(p_{1,i}(x)), x_{\sigma_i(x)}(p_{2,i}(x)) \in W_{\varepsilon, \delta}(x)\) and as \(X_i(x)\) intersect with \(W_{\varepsilon, \delta}(x)\), the definition above is well defined.

We will next increase the number of sets in these coverings, so that they are guaranteed not to fall into any of the “local pieces” at \(x\): although by the definitions (E.3) and (E.6) the local and winding pieces at any \(x\) are disjoint, it may be the case that when we pass to the covering sets \((X_i(x))_{x \in [N(x)]}\), we overlap with the local piece. In particular, consider a “local piece” \(L_{\sqrt{\varepsilon}/\kappa}(x)\) defined as in (E.3), which from the definition does not intersect with \(W_{\varepsilon, \delta}(x)\). For each \(i\), as the boundary points of

\[^{6}\text{Within the lemma, a curve segment means } \{x_{\sigma}(s) \mid s \in [s_1, s_2]\} \subseteq \mathcal{M}_{\sigma} \text{ for certain } \sigma, s_1 \text{ and } s_2 \text{ with } |s_1 - s_2| < \text{len}(\mathcal{M}_{\sigma}), \text{ and we call } |s_1 - s_2| \text{ the length of the curve segment.}\]
We next derive additional properties of the pieces $X_i(x)$ fall in $W_{ε,δ}(x)$, these boundary points do not belong to $L_{\sqrt{E/κ}}(x)$. And as $L_{\sqrt{E/κ}}(x)$ is topologically connected and one dimensional, if $X_i(x)$ intersects with $L_{\sqrt{E/κ}}(x)$, it must contains the whole local piece. As $X_i(x)$ itself is a curve segment and one dimensional, and $L_{\sqrt{E/κ}}(x)$ is open, removing $L_{\sqrt{E/κ}}(x)$ would leave two curve segments with smaller length. Then these two curve segments lie in $M \setminus L_{\sqrt{E/κ}}(x)$, and cover $X_i(x) \setminus L_{\sqrt{E/κ}}(x)$. In other words, for any $x \in M$, there exists $N'(x) \leq \frac{4\sqrt{E/κ}ε}{δ}(M)$ and for $i \in \{1, \cdots, N'(x)\}$, there exist signs $σ_i(x) \in \{\pm\}$ and intervals $I'_i(x) = [s'_{i,0}(x), s'_{i,1}(x)]$ with length no greater than $\frac{1}{\sqrt{1+κ^2}}$ such that

$$W_{ε,δ}(x) \subseteq \bigcup_{i=1}^{N'(x)} X'_i(x),$$

where $X'_i(x) = \{x_{σ_i(x)}(s) \mid s \in I'_i(x)\} \subseteq M \setminus L_{\sqrt{E/κ}}(x)$ with $X'_i(x) \cap W_{ε,δ}(x) \neq \emptyset$. We therefore have

$$\int_{x' \in W_{ε,δ}(x)} ψ^0(\angle_0(x, x')) ds' \leq \sum_{i=1}^{N'(x)} \int_{s \in I'_i(x)} ψ^0(\angle_0(x, x_{σ_i(x)}(s))) ds. \quad (E.40)$$

We next derive additional properties of the pieces $X'_i(x)$ that will allow us to obtain suitable estimates for the integrals on the RHS of $(E.40)$. As each $X'_i(x)$ is a compact set, we let

$$s^*_i(x) \in \arg\min_{s \in I'_i(x)} \angle_0(x, x_{σ_i(x)}(s))$$

and denote $x^*_i(x) = x_{σ_i(x)}(s^*_i(x))$. Below we will abbreviate $x^*_i(x)$, $s^*_i(x)$ and $σ^*_i(x)$ as $x^*_i$, $s^*_i$ and $σ^*_i$ when the base point $x$ is clear. We further abbreviate $x^*_i = x_{σ_i(x)}(s^*_i(x))$. As $X'_i(x)$ intersects with the winding component, we have $\angle_0(x, x^*_i) \leq \frac{2π}{\sqrt{E/κ}} < \frac{π}{2}$. And as $X'_i(x) \cap L_{\sqrt{E/κ}}(x) = \emptyset$, we have $d_M(x, x^*_i) \geq \sqrt{E/κ}$. This means $x^*_i \in W_{ε,δ}(x)$ from $(E.6)$. As cos is strictly decreasing from $0$ to $π$ and $s^*_i$ minimizes $\angle_0(x, x_{σ_i(x)}(s))$, it also maximizes $\langle x, x_{σ_i(x)}(s) \rangle$. For any $s \in I'_i(x)$, from the second order Taylor expansion of $x_{σ_i(x)}(s)$ around $x^*_i$ we have

$$\langle x, x^*_i \rangle \leq \langle x, x_{σ_i(x)}(s) \rangle \leq \langle x, x^*_i \rangle + (s - s^*_i)\langle x, x_{σ_i(x)}(s) \rangle + \left\langle x, \int_{a=s^*_i}^{s} \int_{b=s^*_i}^{a} x_{σ_i(x)}^{(2)}(b) db da \right\rangle \leq \langle x, x^*_i \rangle + (s - s^*_i)^2 M_2,$$

with the last line following from Cauchy-Schwarz. In the previous equations, we are of course using the convention that for a real-valued function $f$ and numbers $a < b$, the notation $\int_a^b f(x) dx$ denotes the integral $\int_a^b f(x) dx$. We are going to use this bound to reprove a classical first-order optimality condition for interval-constrained problems. We split into cases depending on where the point $s^*_i$ lies: if $s^*_i$ is not the right end point $s'_{2,i}$, by taking $s$ approaching $s^*_i$ from above, we would have $\langle x, x^*_i \rangle \leq 0$. Similarly, if $s^*_i$ is not the left end point $s'_{1,i}$, by taking $s$ approaching $s^*_i$ from below, we would have $\langle x, x^*_i \rangle \geq 0$. This gives

$$\langle x, x^*_i \rangle = \begin{cases} \leq 0 & s^*_i = s_{2,i} \\ \geq 0 & s^*_i = s_{1,i} \\ = 0 & o.w. \end{cases}$$

which implies

$$\langle s - s^*_i, x, x^*_i \rangle \leq 0, \quad \forall s \in I'_i(x). \quad (E.41)$$

We use again the Taylor expansion at $s^*_i$ and get

$$\left\| x_{σ_i(x)}(s) - x^*_i - (s - s^*_i)x^*_i \right\|_2 = \left\| \int_{a=s^*_i}^{s} \int_{b=s^*_i}^{a} x_{σ_i(x)}^{(2)}(b) db da \right\|_2.$$
\[ \leq \frac{(s - s_i^*)^2}{2} M_2 \]
\[ = \frac{1}{2} (1 + \kappa^2)^{1/2} (s - s_i^*)^2 \]  
(E.42)

with an application of (E.27) in the last line. Moreover, we have

\[
\|x - x_i^*\|_2 = 2 \sin \left( \frac{\angle(x, x_i^*)}{2} \right) 
\geq \frac{4}{\pi} \sin \left( \frac{\pi}{4} \right) \angle(x, x_i^*) 
= \frac{2\sqrt{2}}{\pi} \angle(x, x_i^*) 
\geq \frac{2\sqrt{2}}{\pi} \Delta_x, 
\]  
(E.43)

where the first line is a trigonometric identity, the first inequality uses \( \angle(x, x_i^*) < \pi/2 \) together with the fact that sin function is concave from 0 to \( \pi \) and thus \( \sin(at) \geq a \sin(t) \) for \( a \in [0, 1] \) and \( t \in [0, \pi] \) (applied to \( a = \angle(x, x_i^*)/(\pi/2) \) and \( t = \pi/4 \)), and the last line follows directly from the definition of \( \Delta_x \) in (C.8). Making use of the preceding estimates, for any \( s \in I_i'(x) \) we can finally calculate

\[
\|x_{\sigma_i}(s) - x\|^2 = \|x_i^* - x + (s - s_i^*) \hat{x}_i^* + (x_{\sigma_i}(s) - x_i^* - (s - s_i^*) \hat{x}_i^*)\|^2 
\geq \|x_i^* - x + (s - s_i^*) \hat{x}_i^*\|^2 - \|x_{\sigma_i}(s) - x_i^* - (s - s_i^*) \hat{x}_i^*\|^2 
\geq \|x_i^* - x\|^2 + \|(s - s_i^*) \hat{x}_i^*\|^2 - 2 \langle x, (s - s_i^*) \hat{x}_i^* \rangle 
= \left( \frac{1}{2} (1 + \kappa^2)^{1/2} (s - s_i^*)^2 \right)^2 
\geq \left( \frac{2\sqrt{2}}{\pi} \Delta_x \right)^2 + (s - s_i^*)^2 - \frac{1}{4} (1 + \kappa^2) (s - s_i^*)^4 
\geq \left( \frac{2\sqrt{2}}{\pi} \Delta_x \right)^2 + \frac{3}{4} (s - s_i^*)^2 
\geq \left( \frac{2}{\pi} \Delta_x + \frac{\sqrt{3}}{2\sqrt{2}} |s - s_i^*| \right)^2. \]  
(E.44)

Above, the second line uses the triangle inequality, the third line uses the parallelogram identity plus Lemma E.3 (first term) and (E.42) (second term), the fourth line comes from (E.43) and (E.41), and the fifth line comes from our construction that the length of each interval \( I_i'(x) \) is no greater than \( 1/\sqrt{1 + \kappa^2} \) and therefore the same is true of \( |s - s_i^*| \). The last line is an application of inequality of arithmetic and geometric means. Additionally, for any \( x, x' \) of unit norm, one has

\[ \angle(x, x') \geq 2 \sin \left( \frac{\angle(x, x')}{2} \right) = \|x - x'\|_2. \]

Combining this and (E.44), for all \( s \in I_i'(x) \) we have

\[ \angle(x_{\sigma_i}(s), x) \geq \|x_{\sigma_i}(s) - x\|_2 \geq \frac{2}{\pi} \Delta_x + \frac{\sqrt{3}}{2\sqrt{2}} |s - s_i^*| \]
\[ \geq \frac{1}{\sqrt{3}} \Delta_x + \frac{1}{\sqrt{3}} |s - s_i^*|, \]

where the last line just worst-cases constants for simplicity. From Lemma G.5, \( \psi^\circ \) is nonnegative and strictly decreasing, so

\[ \int_{s \in I_i'(x)} \psi^\circ \left( \angle(x, x_{\sigma_i}(s)) \right) ds \]
where again, the second to third line comes from the fact that our intervals has length at most $1/\sqrt{1 + \kappa^2}$. From (F.11) in Lemma F.9 and a summation over all $N'(x) \leq 4\Theta_{\epsilon,\delta}(M)$ segments in the covering, there exists constant $C'$ such that when $L \geq C'$,

$$
\sum_{i=1}^{N'(x)} \int_{s \in I_i(x)} \psi^0 \left(\angle (x, x_{\sigma_i}(s))\right) ds \\
\leq \Theta_{\epsilon,\delta}(M) C' n \log \left(\frac{1 + (L - 3) \left(\frac{1}{\sqrt{3}} \Delta \epsilon + \frac{1}{\sqrt{3}} \sqrt{1 + \kappa^2} \right) / (3\pi)}{1 + (L - 3) \frac{\Delta \epsilon}{\sqrt{3}/\kappa} / (3\pi)}\right) \\
\leq \Theta_{\epsilon,\delta}(M) C' n \log \left(1 + \frac{\sqrt{1 + \kappa^2} / \Delta \epsilon}{\kappa}\right).
$$

Recalling our bound (E.40), we can thus take a supremum over $x \in M$ and conclude.

**Lemma E.10.** Let $\epsilon \in (0, 1)$, $\delta \in (0, 1 - \epsilon)$. Let $F_{\epsilon,\delta}$ as in (E.5). There exist constants $C, C'$ such that when $L \geq C$, we have for any $x \in M$,

$$
\int_{x' \in F_{\epsilon,\delta}(x)} \psi^0 \left(\angle (x, x')\right) ds' \leq C' \text{len}(M) n \frac{\kappa}{\delta \sqrt{\epsilon}}.
$$

**Proof.** We have the simple bound from Lemma F.8 and decreasingness of $\psi^0$ from Lemma G.5, that there exists constant $C''$, with

$$
\int_{x' \in F_{\epsilon,\delta}(x)} \psi^0 \left(\angle (x, x')\right) ds' \leq \text{len}(M) \psi^0 \left(\frac{\delta \sqrt{\epsilon}}{\kappa}\right) \\
\leq \text{len}(M) C' n \frac{L - 3}{1 + (L - 3) \frac{\delta \sqrt{\epsilon}}{\kappa} / (3\pi)} \\
\leq \text{len}(M) C' n \frac{\kappa}{\delta \sqrt{\epsilon}},
$$

as claimed.

**E.2 Certificates for the DC-Subtracted Kernel**

**Proof Sketch and Organization** In Appendix E.1, we constructed a certificate for the DC subtracted kernel $\Theta^0$ over the subspace $S_{x}$. In this section, we show that the certificate $g = g_{x [\zeta]}$ defined in Theorem E.1 can also be viewed as the certificate without subspace constraints, satisfying

$$
\Theta^0 [g_{x [\zeta]}] \approx \zeta.
$$
As \( P_{S_\epsilon} \mathcal{O}[g_\epsilon[\zeta]] = \zeta \), we only need \( P_{S_\epsilon} \mathcal{O}[g_\epsilon[\zeta]] \) to be small. The subspace \( S_\epsilon \) is formed by all Fourier basis with low frequency, and thus contains functions that do not oscillate rapidly, in the sense that for any function \( h \) and integer \( k \)

\[
\| P_{S_\epsilon} h \|_{L^2} \lesssim \| \frac{d^k}{dx^k} h \|_{L^2} \dim(S_\epsilon)^{\epsilon}.
\]

This argument is made rigorous in Lemma E.23; by choosing \( k = 3 \) and extracting the dimension of the subspace from (E.12), we obtain the estimate we are looking for. This leaves us to show the derivatives of \( \mathcal{O}[g_\epsilon[\zeta]] \) are small compared to its norm.

The remainder of this subsection is organized as follows. We define a relevant notion of derivatives for the kernel \( \mathcal{O} \) in Definition E.11. These derivatives can be represented as a function of the higher order derivatives of \( \psi \) and that of the angle function (Lemma E.13). We bound the derivatives of the angle by higher order curvatures in Lemmas E.18 to E.20 by geometric parameters of the manifold \( \mathcal{M} \), including higher order regularity constants \( M \) and the angle injectivity radius \( \Delta_\epsilon \).

In Lemma E.22, we show that the projection operator \( P_{S_\epsilon} \) and main invariant operator \( \hat{M}_\epsilon \) commute with differential operators on functions on \( \mathcal{M} \), and thus the “low oscillation” property of the target function \( \zeta \) can be transferred to the “low oscillation” of \( g_\epsilon[\zeta] \) and further down to that of \( \mathcal{O}[g_\epsilon[\zeta]] \).

To simplify the language, we introduce Definition E.24 to represent the required regularity property, and prove that \( g_\epsilon[\zeta] \) and \( \mathcal{O}[g_\epsilon[\zeta]] \) satisfy such regularity in Lemmas E.25 and E.27. Finally, we get control of \( P_{S_\epsilon} \mathcal{O}[g_\epsilon[\zeta]] \) in Lemma E.28.

**Definition E.11.** For any \( x, x' \in \mathcal{M} \), let \( \sigma, \sigma' \in \{\pm 1\} \) denote the class memberships of \( x \) and \( x' \), let \( s, s' \in \mathbb{R} \) be such that \( x_\sigma(s) = x \), \( x_{\sigma'}(s') = x' \), and write \( \mathcal{O}^{(0)}(x, x') = \mathcal{O}(x, x') = \psi(\angle(x, x')) \). We consider higher order derivatives of the kernel with respect to a “simultaneous advance”. For \( i = 1, 2, 3 \), define inductively

\[
\mathcal{O}^{(i)}(x, x') = \frac{d}{dt} \mathcal{O}^{(i-1)}(x_\sigma(s+t), x_{\sigma'}(s'+t)) \bigg|_{x_\sigma(s)=x, x_{\sigma'}(s')=x'}.
\]

Let \( \mathcal{O}^{(i)} \) denote the Fredholm integral operator associated to \( \mathcal{O}^{(i)} \):

\[
\mathcal{O}^{(i)}[h](x) = \int_{x' \in \mathcal{M}} \mathcal{O}^{(i)}(x, x') h(x') dx'.
\]

It is clear that these definitions do not depend on the choice of \( s, s' \in \mathbb{R} \) among ‘equivalent’ points (c.f. (C.5) and surrounding discussion).

**Remark E.12.** For the moment, we have elided the issue that due to differentiability issues with the angle function \( (x, x') \mapsto \angle(x, x') \), the kernels \( \mathcal{O}^{(i)} \) defined in Definition E.11 may not be well-defined on all of \( \mathcal{M} \times \mathcal{M} \). This issue is resolved in Lemma E.13.

**Lemma E.13.** Let

\[
\lambda_0(x, x') = \angle(x, x')
\]

\[
\lambda_{i+1}(x, x') = \frac{d}{dt} \lambda_i(x_\sigma(s+t), x_{\sigma'}(s'+t)) \bigg|_{x_\sigma(s)=x, x_{\sigma'}(s')=x'},
\]

\[
\left[ \left( \frac{\partial}{\partial s} + \frac{\partial}{\partial s'} \right) \right] \lambda_i(x_\sigma(s), x_{\sigma'}(s')) \bigg|_{x_\sigma(s)=x, x_{\sigma'}(s')=x'},
\]

\( i = 0, 1, 2 \),

denote derivatives of the angle function with respect to a “simultaneous advance”. Then when the parameterizations \( x_\sigma \) are five times continuously differentiable (as required in Appendix C.1), these functions are well-defined on \( \mathcal{M} \times \mathcal{M} \).

In addition, the kernels \( \mathcal{O}^{(i)} \) defined in Definition E.11 are well-defined on \( \mathcal{M} \times \mathcal{M} \) and can be expressed in terms of the derivatives of \( \psi \) and the functions \( \lambda_i \) as

\[
\mathcal{O}^{(0)}(x, x') = \psi^c(\angle(x, x'))
\]

30
\[ \Theta^{(1)}(x, x') = \psi(\angle(x, x'))\lambda_1(x, x') \]
\[ \Theta^{(2)}(x, x') = \psi(\angle(x, x'))\lambda_1^2(x, x') + \psi(\angle(x, x'))\lambda_2(x, x') \]
\[ \Theta^{(3)}(x, x') = \psi(\angle(x, x'))\lambda_1^3(x, x') + 3\psi(\angle(x, x'))\lambda_2(x, x')\lambda_1(x, x') \]

where \( \psi, \tilde{\psi}, \tilde{\tilde{\psi}} \) denote the first three derivatives of \( \psi \).

**Proof.** Because the function \( t \mapsto \cos^3(t) \) is infinitely differentiable except at \( \{ -1, 1 \} \subset [-1, 1] \) and \( \psi \) is 3 times continuously differentiable on \([0, \pi]\) (Lemma G.5), and given the differentiability assumption on the curves and the fact that (C.1) precludes \( \mathcal{M} \) from containing any antipodal points, the claim follows immediately by the chain rule except on the diagonal \( \{(x, x) | x \in \mathcal{M}\} \). Here, suppose \( s, s' \) are such that \( x_\sigma(s) = x_\sigma(s') \). Then we have \( x_\sigma(s + t) = x_\sigma(s' + t) \) for every \( t \in \mathbb{R} \). In particular, \( \angle(x_\sigma(s + t), x_\sigma(s' + t)) = 0 \) for all \( t \in \mathbb{R} \), which implies that \( \lambda_i(x, x) = 0 \) for all \( i \). A similar argument implies well-definedness of \( \Theta^{(i)}(x, x) \) for all \( i \), which establishes the claimed formulas on all of \( \mathcal{M} \times \mathcal{M} \).

---

**Lemma E.14.** For points \( x, x' \in \mathcal{M} \) and \( \varepsilon \in (0, 1) \) we have

\[
\sqrt{1 - (x^*x')^2} \geq \begin{cases} \frac{d_\mathcal{M}(x, x')}{s} & d_\mathcal{M}(x, x') \leq \frac{1}{\varepsilon} \\ \frac{2}{\pi} \Delta_x & d_\mathcal{M}(x, x') \geq \frac{2\varepsilon}{\pi} \end{cases}
\]

**Proof.** When \( d_\mathcal{M}(x, x') \geq \frac{2\varepsilon}{\pi} \), from definition of the angle injectivity radius in (C.8) we have \( \angle(x, x') \geq \Delta_x \). From (C.1) we also have \( \angle(x, x') \leq \pi/2 \), then

\[
\sqrt{1 - (x^*x')^2} = \sin(\angle(x, x')) \\
\geq \frac{2}{\pi} \Delta_x,
\]

where the first inequality comes from the monotonicity of \( \sin(t) \) from 0 to \( \pi/2 \). The second inequality uses concavity of \( \sin \) to get \( \sin(t) \geq (2/\pi)t \) for \( 0 \leq t \leq \pi/2 \), and the fact that \( \varepsilon < 1 \) and hence \( \Delta_x \leq \pi/2 \).

When \( d_\mathcal{M}(x, x') \leq \frac{1}{\varepsilon} \), assume \( x, x' \) are parameterized by \( x_\sigma(s), x_\sigma(s') \) separately with \( |s - s'| = d_\mathcal{M}(x, x') \), then \( |s' - s| \leq \frac{1}{\varepsilon} \). Assuming without loss of generality that \( s' \geq s \), using a second-order Taylor expansion and properties from Lemma E.3 gives

\[
x_\sigma(s)^*x_\sigma(s') = x_\sigma(s)^* \left( x_\sigma(s) + (s' - s)x_\sigma(s) + \int_{a=s}^{s'} \int_{b=s}^{a} \tilde{x}_\sigma(b) \, db \, da \right)
\]

\[
= 1 + \int_{a=s}^{s'} \int_{b=s}^{a} \langle x_\sigma(s), \tilde{x}_\sigma(b) \rangle \, db \, da
\]

\[
= 1 + \int_{a=s}^{s'} \int_{b=s}^{a} \langle x_\sigma(b) + (s - b)x_\sigma(s) + \int_{c=b}^{s} \tilde{x}_\sigma(d) \, dd \, dc, \tilde{x}_\sigma(b) \rangle \, db \, da
\]

\[
= 1 - \frac{(s' - s)^2}{2} + \int_{a=s}^{s'} \int_{b=s}^{a} \int_{c=b}^{s} \tilde{x}_\sigma(d)^* \tilde{x}_\sigma(b) \, dd \, dc \, db \, da,
\]

with a Taylor expansion at \( b \) used in the third line, and using the convention that for a real-valued function \( f \) and numbers \( a < b \), the notation \( \int_a^b f(x) \, dx \) denotes the integral \( -\int_b^a f(x) \, dx \). As \( \tilde{k} = \max\{\kappa, \frac{\pi}{2}\} \), we can use the previous expression (with a bound of the integrand in the last line by \( M_\sigma \)) and Lemma E.3 again to obtain after an integration

\[
|x_\sigma(s)^*x_\sigma(s') - 1 + \frac{1}{2}(s' - s)^2| \leq \frac{(s' - s)^4}{4!}(1 + \kappa^2)
\]

31
When specific remainder of the proof, with an abuse of notation we will write \( s \sigma \)

\[
\text{Proof. Let } x, x' \in M. \text{ We have}
\]

\[
| \lambda_1(x, x') | \leq \left\{ \begin{array}{ll}
\frac{7 d_M(x, x')^3}{12} M_4 & d_M(x, x') \leq \frac{1}{\kappa} \\
\frac{1}{\kappa} & \forall x, x' \in M.
\end{array} \right.
\]

\[\text{Proof. Let } s, s' \text{ be such that } x = x_\sigma(s) \text{ and } x' = x_\sigma(s'), \text{ with } |s - s'| = d_M(x, x') \text{ when in addition } \sigma = \sigma'. \text{ As } \angle(x, x') = \cos^{-1}(x^*x'), \]

\[
\lambda_1(x, x') = \frac{\partial}{\partial s} \angle(x_\sigma(s), x_\sigma(s')) + \frac{\partial}{\partial s'} \angle(x_\sigma(s), x_\sigma(s'))
\]

\[
= - \frac{x^*x'}{\sqrt{1 - (x^*x')^2}}.
\]

\[\text{Notice that } \sqrt{1 - (x^*x')^2} = \| (I - xx^*)x' \|_2, \text{ and therefore by Lemma E.3 and Cauchy-Schwarz}
\]

\[
\frac{x^*x'}{\sqrt{1 - (x^*x')^2}} = \left| \frac{\langle x, (I - xx^*)x' \rangle}{\| (I - xx^*)x' \|_2} \right| \leq 1,
\]

and thus \(| \lambda_1(x, x') | \leq 2 \) by symmetry.

When \( d_M(x, x') \leq \frac{1}{\kappa} \), we have \( \sigma = \sigma' \), (as above) \(|s' - s| \leq \frac{1}{\kappa} \). By symmetry, we may assume \( s' \geq s \). From Lemma E.3, we have \( \dot{x}^*x = \dot{x}^*x = 0, \ddot{x}^*x = 1, \dddot{x}^*x = -\frac{1}{3!}x^*(4) \). In the remainder of the proof, with an abuse of notation we will write \( \dot{x} = x_\sigma(s), \ddot{x} = \dot{x} \sigma(s'), \text{ and so on for the higher derivatives to represent the specific points of interest concisely. Thus by a fourth-order Taylor expansion (respectively, of } x' = x_\sigma(s') \text{ at } s, \text{ and of } x = x_\sigma(s) \text{ at } s')
\]

\[
|\dot{x}^*x' + \ddot{x}^*x'|
\]

\[
= \left| \dot{x}^* \left( x + (s' - s) \dot{x} + \frac{(s' - s)^2}{2} \ddot{x} + \frac{(s' - s)^3}{3!} x^{(3)} \right)
\]

\[
+ \int_{a=s}^{s'} \int_{b=s}^{a} \int_{c=s}^{b} \int_{d=s}^{c} x_\sigma^{(4)}(d) dd dc db da \right).
\]

32
\[ + \hat{x}' + \left( x' - (s' - s)x' + \frac{(s' - s)^2}{2} \hat{x}' - \frac{(s' - s)^3}{3!}x' \right) \]

\[ + \int_{a=s}^{s'} \int_{b=a}^{s'} \int_{c=b}^{s'} \int_{d=c}^{s'} x_\sigma^{(4)}(d) dd dc db da \]

\[ = s' - s + \hat{x}' + \frac{(s' - s)^3}{3!} x_\sigma^{(3)} + \int_{a=s}^{s'} \int_{b=a}^{s'} \int_{c=b}^{s'} \int_{d=c}^{s'} x_\sigma^{(4)}(d) dd dc db da \]

\[ - (s' - s) + \hat{x}' - \frac{(s' - s)^3}{3!} x_\sigma^{(3)} + \int_{a=s}^{s'} \int_{b=a}^{s'} \int_{c=b}^{s'} \int_{d=c}^{s'} x_\sigma^{(4)}(d) dd dc db da \]

\[ \leq \frac{(s' - s)^3}{3!} |x_\sigma^{(3)} - \hat{x}'| + \frac{2(s' - s)^4}{4!} M_4 \]

\[ \leq \frac{(s' - s)^3}{3!} \int_{a=s}^{s'} |\hat{x}_\sigma(a)x_\sigma^{(4)}(a) + \hat{x}_\sigma(a)x_\sigma^{(3)}(a)| da + \frac{2(s' - s)^4}{4!} M_4 \]

\[ = \frac{(s' - s)^3}{3!} \int_{a=s}^{s'} \frac{2}{3} |\hat{x}_\sigma(a)x_\sigma^{(4)}(a)| da + \frac{2(s' - s)^4}{4!} M_4 \]

\[ \leq \frac{7}{36} M_4 (s' - s)^4. \quad (E.47) \]

Above, the first inequality uses the triangle inequality and Cauchy-Schwarz; the second inequality Taylor expands the first term in the difference at (s) which leads to a cancellation with the second term and uses the triangle inequality to move the absolute value inside the integral; the following line rewrites using Lemma E.3; and then the final line uses Cauchy-Schwarz, integrates and collects constants. Using Lemma E.14, we obtain that when \( d_M(x, x') \leq \frac{1}{\kappa} \),

\[ |\lambda_1(x, x')| \leq \frac{7d_M^3(x, x')}{12} M_4. \]

\[ \square \]

**Lemma E.16.** There exists an absolute constant \( C \) such that for any \( \varepsilon \in (0, 1) \) and \( x, x' \in M \) we have

\[ |\lambda_2(x, x')| \leq \begin{cases} C(M_4^2 + M_5) d_M(x, x')^3, & d_M(x, x') \leq \frac{1}{\varepsilon^2}, \\ \frac{\varepsilon}{4} \Delta^{-1} + 2M_2, & d_M(x, x') > \frac{1}{\varepsilon^2}. \end{cases} \quad (E.48) \]

**Proof.** Let \( s, s' \) be such that \( x = x_\sigma(s) \) and \( x' = x_{\sigma'}(s') \), with \( |s - s'| = d_M(x, x') \) when in addition \( \sigma = \sigma' \). From (E.46),

\[ \lambda_2(x, x') = \frac{d}{dt} \lambda_1(x_\sigma(s + t), x_{\sigma'}(s' + t)) \]

\[ = -\frac{(\hat{x}' + \hat{x}' x' x') x' x'}{(1 - (x' x')^2)^{3/2}} = \frac{\hat{x}' + \hat{x}' x' x'}{\sqrt{1 - (x' x')^2} x' x'}. \quad (E.49) \]

First consider the case where \( d_M(x, x') > \frac{1}{\varepsilon^2} \). As \( \sqrt{1 - (x' x')^2} = \| (I - xx^*) x'\|_2 \), we can write

\[ \frac{\hat{x}' x'}{\sqrt{1 - (x' x')^2}} = \hat{x}' \| (I - xx^*) x'\|_2 + \frac{\hat{x}' x' x'}{\sqrt{1 - (x' x')^2}} \]

\[ = \hat{x}' \| (I - xx^*) x'\|_2 - \frac{x' x'}{\sqrt{1 - (x' x')^2}} \]

using Lemma E.3. Thus following (E.46) and (E.45) and Lemmas E.3, E.14 and E.15,

\[ |\lambda_2(x, x')| \leq \frac{\hat{x}' x'}{\sqrt{1 - (x' x')^2}} + \frac{|\hat{x}' x' - 2x' x'|}{\sqrt{1 - (x' x')^2}} \]
\[ \frac{(I - x^*)x'}{\| (I - x'^*)x' \|_2} + \frac{(I - x'^*)x}{\| (I - x^*)x \|_2} \leq (4 + 4) \left( \frac{2 \Delta_x}{\pi} \right)^{-1} + 2M_2 \]
\[ = \frac{\pi}{2} \Delta_x^{-1} + 2M_2. \]

When \( d_M(x, x') \leq \frac{1}{4} \), we have \( \sigma = \sigma' \) and \( |s' - s| \leq \frac{1}{4} \). By symmetry, we may assume \( s' \geq s \). Following Lemma E.3, we have \( \dot{x}^* \dot{x} = 0, \dot{x}\dot{x}' = -1 \). In the remainder of the proof, with an abuse of notation we will write \( \dot{x} = \dot{x}_s(s), \dot{x}' = \dot{x}_s(s') \), and so on for the higher derivatives to represent the specific points of interest concisely. We can calculate by Taylor expansion and Lemma E.3

\[ \dot{x}^* \dot{x}' = \ddot{x}^* (x + (s' - s)x + \frac{(s' - s)^2}{2} \ddot{x} + \frac{(s' - s)^3}{6} x^{(3)}) + \int_{a=s}^{s'} \int_{b=s}^{s'} \int_{c=s}^{s'} \int_{d=s}^{s'} x^{(4)} (d) dd dc db da, \]
\[ = -1 + \ddot{x}^* (x + (s' - s)x + \frac{(s' - s)^2}{2} \ddot{x} + \frac{(s' - s)^3}{6} x^{(3)}) + \int_{a=s}^{s'} \int_{b=s}^{s'} \int_{c=s}^{s'} \int_{d=s}^{s'} x^{(4)} (d) dd dc db da, \]
\[ = -1 + \ddot{x}^* (x - (s' - s)x + \frac{(s' - s)^2}{2} \ddot{x}' + \frac{(s' - s)^3}{6} x^{(3)}) + \int_{a=s}^{s'} \int_{b=s}^{s'} \int_{c=s}^{s'} \int_{d=s}^{s'} x^{(4)} (d) dd dc db da, \]
\[ = 1 + \ddot{x}^* (x + (s' - s)x + \frac{(s' - s)^2}{2} \ddot{x} + \frac{(s' - s)^3}{6} x^{(4)}) + \int_{a=s}^{s'} \int_{b=s}^{s'} \int_{c=s}^{s'} \int_{d=s}^{s'} x^{(5)} (d) dd dc db da, \]
\[ = 1 + \ddot{x}^* (x + (s' - s)x + \frac{(s' - s)^2}{2} \ddot{x}' + \frac{(s' - s)^3}{6} x^{(4)}) + \int_{a=s}^{s'} \int_{b=s}^{s'} \int_{c=s}^{s'} \int_{d=s}^{s'} x^{(5)} (d) dd dc db da, \]
\[ = 1 + \ddot{x}^* (x + (s' - s)x + \frac{(s' - s)^2}{2} \ddot{x} + \frac{(s' - s)^3}{6} x^{(4)}) + \int_{a=s}^{s'} \int_{b=s}^{s'} \int_{c=s}^{s'} \int_{d=s}^{s'} x^{(5)} (d) dd dc db da, \]
\[ = 1 + \ddot{x}^* (x - (s' - s)x + \frac{(s' - s)^2}{2} \ddot{x}' + \frac{(s' - s)^3}{6} x^{(4)}) + \int_{a=s}^{s'} \int_{b=s}^{s'} \int_{c=s}^{s'} \int_{d=s}^{s'} x^{(5)} (d) dd dc db da, \]
\[ = 1 + \ddot{x}^* (x - (s' - s)x + \frac{(s' - s)^2}{2} \ddot{x}' + \frac{(s' - s)^3}{6} x^{(4)}) + \int_{a=s}^{s'} \int_{b=s}^{s'} \int_{c=s}^{s'} \int_{d=s}^{s'} x^{(5)} (d) dd dc db da, \]
\[ = 1 + \ddot{x}^* (x - (s' - s)x + \frac{(s' - s)^2}{2} \ddot{x}' + \frac{(s' - s)^3}{6} x^{(4)}) + \int_{a=s}^{s'} \int_{b=s}^{s'} \int_{c=s}^{s'} \int_{d=s}^{s'} x^{(5)} (d) dd dc db da, \]
\[ = 1 + \ddot{x}^* (x - (s' - s)x + \frac{(s' - s)^2}{2} \ddot{x}' + \frac{(s' - s)^3}{6} x^{(4)}) + \int_{a=s}^{s'} \int_{b=s}^{s'} \int_{c=s}^{s'} \int_{d=s}^{s'} x^{(5)} (d) dd dc db da, \]

In addition
\[ |\dot{x}^* x^{(4)} - \dot{x}'^* x'^{(4)}| = \left| - \int_{a=s}^{s'} \ddot{x}^* (a)^x(a) + \ddot{x}'^* (a)^x(a) da \right|, \]
\[ \leq |s' - s| (M_2 M_4 + M_5) \]

by Taylor expansion of the first term in the difference on the LHS at \( s' \). From Lemma E.3, \( \ddot{x}^* x^{(3)} = -\frac{1}{3} \ddot{x} \ddot{x}' + \ddot{x}^* x^{(3)} \). Whence adding (E.50), (E.51), (E.52), (E.53) and applying (E.54) we get
\[ |\ddot{x}^* x' + \ddot{x}'^* x + 2 \ddot{x}^* \ddot{x}'| \leq \frac{(s' - s)^2}{2} \left( \ddot{x}^* x + \ddot{x}'^* x^{(3)} \right) + \frac{(s' - s)^2}{2} \left( \ddot{x}^* x + \ddot{x}'^* x^{(3)} \right) \]
When we have $s, s'$, let Lemma E.17. There exists an absolute constant $d$.

From Lemma E.3, $M_2 \leq M_4$. Plugging (E.55) and (E.47) into the bound (E.49) and using Lemma E.14, we obtain that when $d_M(x, x') \leq \frac{1}{r} \leq \frac{7}{2}$

$$
|\lambda_2(x, x')| \leq \left( \frac{3}{|s' - s|} \right)^3 \left( \frac{7 M_4 |s' - s|^4}{36} \right)^2 + \frac{3}{|s' - s|} \frac{7 (M_2 M_4 + M_5)}{36} |s' - s|^4
$$

$$
= \frac{49}{48} M_4^2 |s' - s|^5 + \frac{7}{12} (M_2 M_4 + M_5) |s' - s|^3
$$

$$
\leq C (M_2^3 + M_3) d_M^3(x, x')
$$

for some absolute constant $C > 0$.

**Lemma E.17.** There exists an absolute constant $C > 0$ such that for any $\varepsilon \in (0, 1)$ and $x, x' \in M$ we have

$$
|\lambda_3(x, x')| \leq \begin{cases}
C(M_2^2 + M_3 M_4) d_M(x, x')^3, & d_M(x, x') \leq \frac{1}{r}

\frac{3\pi^2 \Delta \varepsilon^2}{3} + 9\pi M_2 \Delta \varepsilon^{-1} + 2M_3 + 8, & d_M(x, x') > \frac{\sqrt{\varepsilon}}{r}
\end{cases}
$$

**Proof.**

Let $s, s'$ be such that $x = x_\sigma(s)$ and $x' = x_\sigma'(s')$, with $|s - s'| = d_M(x, x')$ when in addition $\sigma = \sigma'$. Then from (E.49),

$$
\lambda_3(x, x') = \frac{d}{dt} \left| \lambda^2(x_\sigma(s + t), x_\sigma'(s' + t)) \right|
$$

$$
= \frac{(\dot{x}^* x + \dot{x}^* x)^3}{(1 - (x^* x')^2)^{3/2}} - \frac{3(\dot{x}^* x + \dot{x}^* x)^2 (x^* x')^2}{(1 - (x^* x')^2)^{5/2}}
$$

$$
- \frac{3(\dot{x}^* x + \dot{x}^* x) x^* x'}{(1 - (x^* x')^2)^{3/2}}
$$

$$
- \frac{3\dot{x}^* \ddot{x} + 3\dot{x}^* \ddot{x} + x^3 x' + x'(3)^x x}{\sqrt{1 - (x^* x')^2}}.
$$

When $d_M(x, x') > \frac{\sqrt{\varepsilon}}{r}$, as $\sqrt{1 - (x^* x')^2} = \|(I - xx^*) x'\|_2$ and from Lemma E.3 $x^{(3)^*} x = 0$,

$$
\frac{x^{(3)^*} x'}{\sqrt{1 - (x^* x')^2}} = \left\langle x^{(3)}, \frac{(I - xx^*) x'}{\|(I - xx^*) x'\|_2} \right\rangle.
$$

Thus from (E.46), (E.49), (E.45), Lemma E.14, Lemma E.15 and Lemma E.16,

$$
|\lambda_3(x, x')| \leq \begin{cases}
|\lambda_1(x, x')^3| + \frac{3\lambda_1(x, x') \lambda_2(x, x') - x^* x'}{\sqrt{1 - (x^* x')^2}} + \frac{3\dot{x}^* \ddot{x} + 3\dot{x}^* \ddot{x}'}{\sqrt{1 - (x^* x')^2}}
\end{cases}
$$

$$
+ \left| \left\langle x^{(3)}, \frac{(I - xx^*) x'}{\|(I - xx^*) x'\|_2} \right\rangle \right| + \left| \left\langle x^{(3)}, \frac{(I - xx^*) x'}{\|(I - xx^*) x'\|_2} \right\rangle \right|
$$

$$
\leq 8 + 6 \left( \frac{7}{4} \frac{\pi}{\Delta \varepsilon^{-1}} + 2M_2 \right) \left( \frac{2}{\pi} \Delta \varepsilon \right)^{-1} + 6M_2 \left( \frac{2}{\pi} \Delta \varepsilon \right)^{-1} + 2M_3
$$

$$
\leq \frac{3\pi^2}{4} \Delta \varepsilon^{-2} + 9\pi M_2 \Delta \varepsilon^{-1} + 2M_3 + 8.
$$

35
When \( d_M(x, x') \leq \frac{1}{s} \), we have \( \sigma = \sigma' \) and \( |s' - s| \leq \frac{1}{s} \). By symmetry, we may assume \( s' \geq s \).

Following Lemma E.3, \( \dot{x}^* \dot{x} = 0 \) and \( \ddot{x}^* \dot{x} = -\dddot{x}^* x^{(3)} \). In the remainder of the proof, with an abuse of notation we will write \( \dot{x} = \dot{x}_\sigma(s) \), \( \ddot{x}' = \dot{x}_\sigma(s') \), and so on for the higher derivatives to represent the specific points of interest concisely. Because we can reuse bounds for lower-order \( \lambda_i \) terms to bound the first three terms in (E.56), we will focus on controlling the last term. We can calculate by Taylor expansion

\[
3 \dddot{x}^* \dddot{x}' + 3 \dddot{x}^* \dddot{x}' + x^{(3)} x' + x^{(3)} x'
\]

\[
= 3 \dddot{x}^* \left( \dddot{x} + (s' - s) \dddot{x} + \frac{(s' - s)^2}{2} x^{(3)} + \frac{(s' - s)^3}{6} x^{(4)} \right)
\]

\[
+ \int_{a=s}^{s'} \int_{b=b}^{b} \int_{c=c}^{c} \int_{d=d}^{d} \dddot{x}_\sigma^{(5)}(d) dd dc db da
\]

\[
+ 3 \dddot{x}^* \left( \dddot{x} - (s' - s) \dddot{x} + \frac{(s' - s)^2}{2} x^{(3)} - \frac{(s' - s)^3}{6} x^{(4)} \right)
\]

\[
+ \int_{a=s}^{s'} \int_{b=b}^{b} \int_{c=c}^{c} \int_{d=d}^{d} \dddot{x}_\sigma^{(5)}(d) dd dc db da
\]

\[
+ x^{(3)} \left( \dddot{x} + (s' - s) \dddot{x} + \frac{(s' - s)^2}{2} x^{(3)} + \frac{(s' - s)^3}{6} x^{(4)} \right)
\]

\[
+ \int_{a=s}^{s'} \int_{b=b}^{b} \int_{c=c}^{c} \int_{d=d}^{d} \dddot{x}_\sigma^{(5)}(d) dd dc db da
\]

\[
+ x^{(3)} \left( \dddot{x} - (s' - s) \dddot{x} + \frac{(s' - s)^2}{2} x^{(3)} - \frac{(s' - s)^3}{6} x^{(4)} \right)
\]

\[
+ \int_{a=s}^{s'} \int_{b=b}^{b} \int_{c=c}^{c} \int_{d=d}^{d} \dddot{x}_\sigma^{(5)}(d) dd dc db da
\]

\[
= 2(s' - s) (\dddot{x}^* \dddot{x} - \dddot{x}'^* \dddot{x}') + 2(s' - s)^2 \left( \dddot{x}^* x^{(3)} + \dddot{x}'^* x^{(3)} \right)
\]

\[
+ \frac{(s' - s)^3}{6} \left( 3 \dddot{x}^* x^{(4)} - 3 \dddot{x}'^* x^{(4)} + x^{(3)} x^{(3)} - x^{(3)} x^{(3)} \right)
\]

\[
+ \int_{a=s}^{s'} \int_{b=b}^{b} \int_{c=c}^{c} \int_{d=d}^{d} \dddot{x}_\sigma^{(5)}(d) dd dc db da
\]

\[
+ \int_{a=s}^{s'} \int_{b=b}^{b} \int_{c=c}^{c} \int_{d=d}^{d} \dddot{x}_\sigma^{(5)}(d) dd dc db da
\]

\[
+ x^{(3)} \int_{a=s}^{s'} \int_{b=b}^{b} \int_{c=c}^{c} \int_{d=d}^{d} \dddot{x}_\sigma^{(4)}(d) dd dc db da
\]

\[
+ x^{(3)} \int_{a=s}^{s'} \int_{b=b}^{b} \int_{c=c}^{c} \int_{d=d}^{d} \dddot{x}_\sigma^{(4)}(d) dd dc db da
\]
We use Taylor expansion again for the second term of (E.58)

\[ - \int_{a=s}^{a=s'} \int_{b=s}^{b=b} \left( \ddot{x}_\sigma(b) \cdot \dddot{x}_\sigma^{(4)}(b) + \dddot{x}_\sigma^{(3)}(b) \cdot \dddot{x}_\sigma^{(3)}(b) \right) \, db \, da. \]

\[ = \int_{b=s}^{b=b} \int_{a=s}^{a=a} \left( \ddot{x}_\sigma(b) \cdot \dddot{x}_\sigma^{(4)}(b) + \dddot{x}_\sigma^{(3)}(b) \cdot \dddot{x}_\sigma^{(3)}(b) \right) \, da \, db \\
- \int_{b=s}^{b=b} \int_{a=s}^{a=a} \left( \ddot{x}_\sigma(b) \cdot \dddot{x}_\sigma^{(4)}(b) + \dddot{x}_\sigma^{(3)}(b) \cdot \dddot{x}_\sigma^{(3)}(b) \right) \, da \, db \\
= \int_{b=s}^{b=b} \int_{a=s}^{a=a} \left( \ddot{x}_\sigma(s) \cdot \dddot{x}_\sigma^{(4)}(s) + \dddot{x}_\sigma^{(3)}(s) \cdot \dddot{x}_\sigma^{(3)}(s) \right) \, db \, da \\
- \frac{1}{2} \left( \ddot{x}_\sigma(b) \cdot \dddot{x}_\sigma^{(4)}(b) + \dddot{x}_\sigma^{(3)}(b) \cdot \dddot{x}_\sigma^{(3)}(b) \right) \, db \\
= \int_{b=s}^{b=b} \int_{a=s}^{a=a} \left( \ddot{x}_\sigma(s) \cdot \dddot{x}_\sigma^{(4)}(s) + \dddot{x}_\sigma^{(3)}(s) \cdot \dddot{x}_\sigma^{(3)}(s) \right) \, db \, da \\
\leq \int_{b=s}^{b=b} \int_{a=s}^{a=a} \left( 2M_2 M_5 + 3M_3 M_4 \right) \, db \, da \\
= (s' - s)^3 \left( 2M_2 M_5 + 3M_3 M_4 \right) \int_{b=s}^{(s+s')/2} \frac{1}{6}. \tag{E.59} \]

We use Taylor expansion again for the second term of (E.58)

\[ 3 \dddot{x}_\sigma \dddot{x}_\sigma^{(4)} - 3 \dddot{x}_\sigma^{(3)} \dddot{x}_\sigma^{(3)} + \dddot{x}_\sigma^{(3)} \dddot{x}_\sigma^{(3)} \dddot{x}_\sigma^{(3)} = - \int_{a=s}^{a=s'} \left[ 3 \dddot{x}_\sigma(a) \cdot \dddot{x}_\sigma^{(5)}(a) + 5 \dddot{x}_\sigma^{(3)}(a) \dddot{x}_\sigma^{(4)}(a) \right] \, da. \tag{E.60} \]

Plug (E.59) and (E.60) back to (E.58) and we conclude that

\[ |3 \dddot{x}_\sigma \dddot{x}_\sigma^{(4)} + 3 \dddot{x}_\sigma^{(3)} \dddot{x}_\sigma^{(3)} + \dddot{x}_\sigma^{(3)} \dddot{x}_\sigma^{(3)} \dddot{x}_\sigma^{(3)}| \leq \frac{(s'-s)^4}{3} \left( M_2 M_5 + 3M_3 M_4 \right) + \frac{(s'-s)^4}{6} \left( 3M_2 M_5 + 5M_3 M_4 \right) \]

\[ + \frac{(s'-s)^4}{4!} \left( 6M_2 M_5 + 2M_3 M_4 \right) \]

\[ = \frac{(s'-s)^4}{12} \left( 13M_2 M_5 + 23M_3 M_4 \right). \tag{E.61} \]
As from Lemma E.3 $M_2 \leq M_4$, when $|s' - s| \leq \frac{1}{\kappa}$, plugging (E.47), (E.55), (E.61), and Lemma E.14 into (E.56), we have

$$
\lambda_3(x, x') \leq \left( \frac{|s' - s|}{3} \right)^{-3} \left( \frac{7}{36} M_4 |s' - s|^4 \right)^3 + 3 \left( \frac{|s' - s|}{3} \right)^{-5} \left( \frac{7}{36} M_4 |s' - s|^4 \right)^3
+ 3 \left( \frac{|s' - s|}{3} \right)^{-3} \left( \frac{7}{36} M_4 |s' - s|^4 \right) \left( \frac{7}{36} (M_2 M_4 + M_5) \right)
+ \left( \frac{|s' - s|}{3} \right)^{-1} |s' - s|^4 \left( 13 M_2 M_5 + 23 M_3 M_4 \right)
= \left( \frac{343}{1728} |s' - s|^9 + \frac{343}{64} |s' - s|^7 M_4^3 + \frac{49}{16} M_4 (M_2 M_4 + M_5) |s' - s|^5 \right)
+ \left( \frac{1}{4} |s' - s|^3 (13 M_2 M_5 + 23 M_3 M_4) \right)
\lesssim (M_4^3 + M_4 M_5 + M_3 M_4) d_{M_4}^3(x, x'),
$$

where the last line uses the fact that we can adjust constants to keep only the lowest-order term involving the distance, given that the distance is bounded.

\[\Box\]

**Lemma E.18.** For $\varepsilon \in (0, \frac{3}{7})$, there exist positive constants $C, C_1$ such that when $L \geq C$, we have

$$
\| \Theta^{(1)} \|_{L^2 \to L^2} \leq P_1(M_4 \text{, len}(\mathcal{M}), \Delta_\varepsilon^{-1}) n,
$$

where $P_1(M_4 \text{, len}(\mathcal{M}), \Delta_\varepsilon^{-1}) = C_1 \left( M_4 + \text{len}(\mathcal{M}) \Delta_\varepsilon^{-2} \right)$ is a polynomial in $M_4 \text{, len}(\mathcal{M})$ and $\Delta_\varepsilon^{-1}$.

**Proof.** From Lemma E.5 and Lemma E.13 we have

$$
\| \Theta^{(1)} \|_{L^2 \to L^2} \leq \sup_{x, x' \in \mathcal{M}} \int_{\mathcal{M}} |\Theta^{(1)}(x, x')| \, dx'
= \sup_{x \in \mathcal{M}} \int_{x' \in \mathcal{M}} |\hat{\psi}(\angle(x, x'))| \lambda_3(x, x') \, dx'
\leq \sup_{x \in \mathcal{M}} \int_{x' \in \mathcal{M}} |\hat{\psi}(\angle(x, x'))| \lambda_1(x, x') \, dx'
$$

Lemmas E.15 and F.10 provide us the control for $\hat{\psi}$ and $\lambda_1$. From Lemma F.10, there exist constants $C_1, C_2$, such that when $L > C$, we have

$$
\max_{t \geq r} |\hat{\psi}(t)| \leq \frac{C_1 n}{r^2}.
$$

From Lemma E.15, we have

$$
|\lambda_1(x, x')| \leq \left( \frac{7 d_{M_4}(x, x')^3}{12} \right) M_4 \quad \forall x, x' \in \mathcal{M}.
$$

In order to get a lower bound for the angle $\angle(x, x')$, for a fixed point $x \in \mathcal{M}$, we decompose the integral into nearby piece $\mathcal{N}(x) = \{ x' \in \mathcal{M} \mid d_{\mathcal{M}}(x, x') \leq \frac{V_2}{\kappa} \}$ and faraway piece $\mathcal{F}(x) = \{ x' \in \mathcal{M} \mid d_{\mathcal{M}}(x, x') > \frac{V_2}{\kappa} \}$ and have

$$
\| \Theta^{(1)} \|_{L^2 \to L^2} \leq \sup_{x \in \mathcal{M}} \left( \int_{x' \in \mathcal{N}(x)} |\hat{\psi}(\angle(x, x'))| \lambda_1(x, x') \, dx' \right.
+ \left. \int_{x' \in \mathcal{F}(x)} |\hat{\psi}(\angle(x, x'))| \lambda_1(x, x') \, dx' \right) \tag{E.62}
$$

Then for any point $x'$ in faraway piece $\mathcal{F}(x)$, $d_{\mathcal{M}}(x, x') > \frac{V_2}{\kappa}$, we have $\angle(x, x') \geq \Delta_\varepsilon$ from (C.8) with $\Delta_\varepsilon \leq \frac{V_2}{\kappa} \leq \frac{\pi}{2}$. From Lemmas E.15 and F.10 we get

$$
\int_{x' \in \mathcal{F}(x)} |\hat{\psi}(\angle(x, x'))| \lambda_1(x, x') \, dx' \leq \int_{x' \in \mathcal{F}(x)} \frac{C_1 n}{\Delta_\varepsilon} \lambda_1(x, x') \, dx'
$$

38
To utilize the bound above, we need to get a lower bound for the angle \( \angle C \) that when Lemmas F.10 and F.11 provide bounds for derivatives of \( M \). From Lemma E.19. we get

\[
\int_{x' \in N(x)} |\psi(\angle(x, x'))| |\lambda_1(x, x')| dx' \leq \int_{s'=s-\frac{\Delta}{k}}^{s+\frac{\Delta}{k}} |\psi(\angle(x, x_{\sigma}(s')))||\lambda_1(x, x_{\sigma}(s'))| ds' \\
\leq \int_{s'=s-\frac{\Delta}{k}}^{s+\frac{\Delta}{k}} \frac{C_1 n}{(1-\varepsilon)^2 |s'-s|^2} |\lambda_1(x, x_{\sigma}(s'))| ds' \\
\leq \int_{s'=s-\frac{\Delta}{k}}^{s+\frac{\Delta}{k}} \frac{C_1 n}{(1-\varepsilon)^2 |s'-s|^2} \frac{7|s'-s|^3}{12} M_4 ds' \\
\leq C'' M_4 n \int_{s'=s-\frac{\Delta}{k}}^{s+\frac{\Delta}{k}} |s'-s| ds' \\
\leq C'' M_4 n \left( \frac{\sqrt{\varepsilon}}{k} \right)^2 \\
\leq C'' M_4 n 
\]

for some constant \( C'' \), where the fourth line comes from \( \varepsilon < \frac{3}{4} \) and the last line comes from \( k \geq \frac{2}{\varepsilon} \) from definition. Plugging (E.63) and (E.64) back in (E.62) proves the claim.

**Lemma E.19.** For \( \varepsilon \in (0, \frac{1}{4}) \), there exist positive constants \( C, C_2 \) such that when \( L \geq C \), we have

\[
\| \Theta^{(2)} \|_{L^2 \to L^2} \leq P_2(M_2, M_4, M_5, \text{len}(M), \Delta^{-1}_e) n
\]

where

\[
P_2(M_2, M_4, M_5, \text{len}(M), \Delta^{-1}_e) = C_2 \left( M_2^2 + M_5 + \text{len}(M) \left( \Delta^{-3}_e + M_2 \Delta^{-2}_e \right) \right)
\]

is a polynomial in \( M_2, M_4, M_5, \text{len}(M) \) and \( \Delta^{-1}_e \).

**Proof.** From Lemma E.5 we have

\[
\| \Theta^{(2)} \|_{L^2 \to L^2} = \sup_{x \in M} \int_{x' \in M} |\Theta^{(2)}(x, x')| dx'.
\]

From Lemma E.13, we know

\[
|\Theta^{(2)}(x, x')| = |\psi(\angle(x, x'))|\lambda_2^2(x, x') + |\psi(\angle(x, x'))|\lambda_2(x, x') \\
\leq |\psi(\angle(x, x'))|\lambda_2^2(x, x') + |\psi(\angle(x, x'))|\lambda_2(x, x').
\]

Lemmas F.10 and F.11 provide bounds for derivatives of \( \psi(t) \): there exist constants \( C_1, C_2 \), such that when \( L > C \), we have

\[
\max_{t \geq r} |\psi(t)| \leq \frac{C_1 n}{r^2}
\]

and

\[
\max_{t \geq r} |\psi(t)| \leq \frac{C_2 n}{r^3}.
\]

To utilize the bound above, we need to get a lower bound for the angle \( \angle(x, x') \). For a fixed point \( x \in M \), we decompose the integral into nearby piece \( N(x) = \{ x' \in M \mid d_M(x, x') \leq \frac{\sqrt{\varepsilon}}{k} \} \) and
faraway piece $\mathcal{F}(x) = \{x' \in \mathcal{M} \mid d_{\mathcal{M}}(x, x') > \frac{\sqrt{3}}{2}\}$. Then for $x' \in \mathcal{F}(x)$, $d_{\mathcal{M}}(x, x') > \frac{\sqrt{3}}{2}$, and we have $\angle(x, x') \geq \Delta_\varepsilon$ with $\Delta_\varepsilon \leq \frac{\sqrt{3}}{2} \leq \frac{\pi}{2}$. From Lemmas E.15 and E.16 we get

$$
\int_{x' \in \mathcal{F}(x)} |\Theta^{(2)}(x, x')|dx' \leq \int_{x' \in \mathcal{F}(x)} \frac{C_2 n}{\Delta_\varepsilon} \lambda_1^2(x, x') + \frac{C_1 n}{\Delta_\varepsilon} \lambda_2(x, x')|dx' \\
\leq \text{len}(\mathcal{M}) \left( \frac{4 C_2 n}{\Delta_\varepsilon} + \frac{C_1 n}{\Delta_\varepsilon} \left( \frac{\pi}{4} \Delta^{-1}_\varepsilon + 2 M_2 \right) \right) \\
\leq C' \left( \Delta_\varepsilon^{-3} + M_2 \Delta_\varepsilon^{-2} \right) \text{len}(\mathcal{M}) n \tag{E.65}
$$

for some constant $C'$.

For nearby piece, let $s, b$ be such that $x = x_{s}(s)$. Follow Lemma E.4 we have $\angle(x, x_{s}(s')) \geq (1 - \varepsilon)|s - s'|$ when $|s - s'| \leq \frac{\sqrt{3}}{2}$. From Lemmas E.15, E.16, F.10 and F.11 there exists constant $c, C''$ such that

$$
\int_{x' \in \mathcal{N}(x)} |\Theta^{(2)}(x, x')|dx' \\
\leq \int_{s' = s - \frac{\sqrt{3}}{2}}^{s + \frac{\sqrt{3}}{2}} |\Theta^{(2)}(x, x_{s}(s'))|ds' \\
\leq \int_{s' = s - \frac{\sqrt{3}}{2}}^{s + \frac{\sqrt{3}}{2}} \frac{C_2 n}{(1 - \varepsilon)^3 |s' - s|^{3}} |\lambda_1^2(x, x_{s}(s'))| + \frac{C_1 n}{(1 - \varepsilon)^2 |s' - s|^{2}} |\lambda_2(x, x_{s}(s'))|ds' \\
\leq \int_{s' = s - \frac{\sqrt{3}}{2}}^{s + \frac{\sqrt{3}}{2}} \frac{C_2 n}{(1 - \varepsilon)^3 |s' - s|^{3}} \left| \frac{2}{12} M_2 \right|^2 + \frac{C_1 n}{(1 - \varepsilon)^2 |s' - s|^{2}} |c(M_2^2 + M_5)| |s' - s|^{3} |ds' \\
\leq C'' \left( 1 + \frac{\sqrt{3}}{\Delta_\varepsilon} \right)^4 \left( M_2^2 + M_5 \right) n \\
\leq C'' \left( M_2^2 + M_5 \right) n. \tag{E.66}
$$

combining (E.65) and (E.66) directly proves the claim. □

**Lemma E.20.** For $\varepsilon \in (0, \frac{1}{3})$, there exist positive constants $C, C_3$ such that when $L \geq C$, we have

$$
\|\Theta^{(3)}\|_{L^2 \rightarrow L^2} \leq P_3(M_2, M_3, M_4, M_5, \text{len}(\mathcal{M}), \Delta_\varepsilon^{-1}) n
$$

where

$$
P_3(M_2, M_3, M_4, M_5, \text{len}(\mathcal{M}), \Delta_\varepsilon^{-1}) = C_3 \left( M_2^3 + M_3 M_4 + M_4 M_5 + \text{len}(\mathcal{M}) \left( \Delta_\varepsilon^{-4} + M_2 \Delta_\varepsilon^{-3} + M_4 \Delta_\varepsilon^{-2} \right) \right)
$$

is a polynomial in $M_2, M_3, M_4, M_5, \text{len}(\mathcal{M})$ and $\Delta_\varepsilon^{-1}$.

**Proof.** From Lemma E.5 we have

$$
\|\Theta^{(3)}\|_{L^2 \rightarrow L^2} = \sup_{x \in \mathcal{M}} \int_{x' \in \mathcal{M}} |\Theta^{(3)}(x, x')|dx'.
$$

From Lemma E.13, we know

$$
|\Theta^{(3)}(x, x')| \\
= |\dot{\psi}(\angle(x, x'))|\lambda_1^3(x, x') + 3\dot{\psi}(\angle(x, x'))|\lambda_2(x, x')|\lambda_1(x, x') + \dot{\psi}(\angle(x, x'))|\lambda_3(x, x')| \\
\leq |\dot{\psi}(\angle(x, x'))|\lambda_1^3(x, x') + 3|\dot{\psi}(\angle(x, x'))|\lambda_2(x, x')|\lambda_1(x, x')| + |\dot{\psi}(\angle(x, x'))|\lambda_3(x, x')|
$$

From Lemmas F.10 to F.12, there exist constants $C_1, C_2, C_3$, such that when $L > C$, we have

$$
\max_{t \geq r} |\dot{\psi}(t)| \leq \frac{C_1 n}{r^2}
$$

where
where the last line comes from the fact that \( \angle(x, x') \), for a fixed point \( x \in M \), we decompose the integral into a nearby piece \( N(x) = \{ x' \in M \mid d_M(x, x') \leq \frac{\sqrt{\epsilon}}{8} \} \) and a faraway piece \( F(x) = \{ x' \in M \mid d_M(x, x') > \frac{\sqrt{\epsilon}}{8} \} \). When \( d_M(x, x') > \frac{\sqrt{\epsilon}}{8} \), we have \( \angle(x, x') \geq \Delta_\varepsilon \) with \( \Delta_\varepsilon \leq \frac{\sqrt{\epsilon}}{\kappa} \leq \frac{\pi}{2} \).

From Lemmas E.15 to E.17, there exist constants \( c, C' \) such that

\[
\int_{x' \in F(x)} |\Theta^{(3)}(x, x')| \, dx' 
\leq \int_{x' \in F(x)} \frac{C_3 n}{\Delta_\varepsilon} \left| \lambda_3^3(x, x') \right| + \frac{C_2 n}{\Delta_\varepsilon^2} \left| \lambda_2(x, x') \right| \left| \lambda_1(x, x') \right| + \frac{C_1 n}{\Delta_\varepsilon^3} \left| \lambda_3(x, x') \right| \, dx'
\leq \text{len}(M) \left( \frac{8C_3 n}{\Delta_\varepsilon} + \frac{2C_2 n}{\Delta_\varepsilon^2} \left( \frac{\pi}{4} \Delta_\varepsilon^{-1} + 2M_2 \right) 
+ \frac{C_1 n}{\Delta_\varepsilon^3} \left( \frac{3\pi^2}{4} \Delta_\varepsilon^{-2} + 9\pi M_2 \Delta_\varepsilon^{-1} + 2M_3 + 8 \right) \right)
\leq C' \left( \Delta_\varepsilon^{-4} + M_2 \Delta_\varepsilon^{-3} + M_3 \Delta_\varepsilon^{-2} \right) \text{len}(M)n
\]

for some constant \( C' \).

For the nearby piece, let \( \sigma, s \) be such that \( x = x_\sigma(s) \). From Lemma E.4, when \( |s - s'| \leq \frac{\sqrt{\epsilon}}{8} \), we have \( \angle(x, x_\sigma(s')) \geq (1 - \varepsilon)|s - s'| \). From Lemmas E.15 to E.17 there exist constants \( c', c'', C'' \) such that

\[
\int_{x' \in N(x)} |\Theta^{(3)}(x, x')| \, dx' \leq \int_{s' = s - \frac{\sqrt{\epsilon}}{8}}^{s - \frac{\sqrt{\epsilon}}{8}} |\Theta^{(3)}(x, x_\sigma(s'))| \, ds'
\leq \int_{s' = s - \frac{\sqrt{\epsilon}}{8}}^{s - \frac{\sqrt{\epsilon}}{8}} \frac{C_3 n}{(1 - \varepsilon)^4|s' - s|^4} \left| \lambda_3^3(x, x_\sigma(s')) \right|
\leq \frac{C_3 n}{(1 - \varepsilon)^4|s' - s|^4} \left| \lambda_2(x, x_\sigma(s')) \right| \left| \lambda_1(x, x_\sigma(s')) \right|
\leq \frac{C_3 n}{(1 - \varepsilon)^4|s' - s|^4} \left| \lambda_3(x, x_\sigma(s')) \right| \, ds'
\leq \int_{s' = s - \frac{\sqrt{\epsilon}}{8}}^{s - \frac{\sqrt{\epsilon}}{8}} \frac{C_2 n}{(1 - \varepsilon)^4|s' - s|^4} \left| \lambda_2^2(x, x_\sigma(s')) \right| \left| \lambda_1^2(x, x_\sigma(s')) \right|
\leq \int_{s' = s - \frac{\sqrt{\epsilon}}{8}}^{s - \frac{\sqrt{\epsilon}}{8}} \frac{C_2 n}{(1 - \varepsilon)^4|s' - s|^4} \left| \lambda_2^2(x, x_\sigma(s')) \right| \left| \lambda_1^2(x, x_\sigma(s')) \right|
\leq C'' \left( M_3^3 + M_3 M_4 + M_4 M_5 \right) |s - s'| \, ds'
\]

where the last line comes from the fact that \( \int_{s' = s - \frac{\sqrt{\epsilon}}{8}}^{s - \frac{\sqrt{\epsilon}}{8}} |s - s'| \, ds' < 2(k^{-1} - 1) \leq 2(2/\pi)^{-1} \). Combining (E.67) and (E.68) directly proves the claim.

**Lemma E.21.** For \( i = 0, 1, 2 \) and any differentiable \( h : M \to \mathbb{C} \), we have

\[
\frac{d}{ds} \Theta^{(i)}[h](x) = \Theta^{(i)} \left[ \frac{d}{ds} h \right](x) + \Theta^{(i+1)}[h](x),
\]

where we recall the notation defined in (C.5).
Proof. Let \( s \) be such that \( x_\sigma(s) = x \). We have

\[
\frac{d}{ds} \Theta^{(i)}[h(x_\sigma(s))] = \frac{\partial}{\partial t} \int_{s' \in \mathcal{M}} \Theta^{(i)}(x_\sigma(s + t), x_{\sigma'}(s')) h(x_{\sigma'}(s')) ds'
\]

\[
= \frac{\partial}{\partial t} \sum_{\sigma'} \int_{s'} \Theta^{(i)}(x_\sigma(s + t), x_{\sigma'}(s')) h(x_{\sigma'}(s')) ds'
\]

\[
= \sum_{\sigma'} \lim_{t \to 0} \frac{1}{t} \left[ \int_{s'} \Theta^{(i)}(x_\sigma(s + t), x_{\sigma'}(s')) h(x_{\sigma'}(s')) ds' \right.
\]

\[
- \left. \int_{s'} \Theta^{(i)}(x_\sigma(s), x_{\sigma'}(s')) h(x_{\sigma'}(s')) ds' \right]
\]

\[
= \sum_{\sigma'} \lim_{t \to 0} \frac{1}{t} \left[ \int_{s'} \Theta^{(i)}(x_\sigma(s + t), x_{\sigma'}(s' + t)) h(x_{\sigma'}(s' + t)) ds'
\]

\[
- \int_{s'} \Theta^{(i)}(x_\sigma(s), x_{\sigma'}(s')) h(x_{\sigma'}(s')) ds'
\]

\[
+ \int_{s'} \Theta^{(i)}(x_\sigma(s), x_{\sigma'}(s')) h(x_{\sigma'}(s')) ds'
\]

\[
- \int_{s'} \Theta^{(i)}(x_\sigma(s), x_{\sigma'}(s')) h(x_{\sigma'}(s')) ds'
\]

\[
= \sum_{\sigma'} \lim_{t \to 0} \frac{1}{t} \left[ \int_{s'} \left[ \Theta^{(i)}(x_\sigma(s + t), x_{\sigma'}(s' + t)) - \Theta^{(i)}(x_\sigma(s), x_{\sigma'}(s')) \right] h(x_{\sigma'}(s' + t)) ds'
\]

\[
\int_{s'} \Theta^{(i)}(x_\sigma(s), x_{\sigma'}(s')) [h(x_{\sigma'}(s' + t)) - h(x_{\sigma'}(s'))] ds'
\]

Above, the domain of each of the \( s' \) integrals is a fundamental domain for the circles \( \mathbb{R}/(\text{len} \mathcal{M}_\sigma) \cdot \mathbb{Z} \) (by periodicity of the parameterizations, the specific fundamental domain is irrelevant). For \( i = 0, 1, 2 \) we have by the mean value theorem

\[
\left| \frac{\Theta^{(i)}(x_\sigma(s + t), x_{\sigma'}(s' + t)) - \Theta^{(i)}(x_\sigma(s), x_{\sigma'}(s'))}{t} \right|
\]

\[
= \left| \Theta^{(i+1)}(x_\sigma(s + t_0), x_{\sigma'}(s' + t_0)) h(x_{\sigma'}(s' + t)) \right|
\]

\[
\leq \sup_{x_1, x_2 \in \mathcal{M}} \left| \Theta^{(i+1)}(x_1, x_2) \right| \sup_{x_3 \in \mathcal{M}} \| h(x_3) \|
\]

for some \( |t_0| \leq |t| \). As \( \mathcal{M} \) is closed with bounded length and \( h \) is differentiable, \( \sup_{x_3 \in \mathcal{M}} h(x_3) \) is bounded. By the formulas in Lemma E.13, the fact that \( \psi \) is \( C^3 \) by Lemma G.5, and Lemmas E.15 to E.17, it follows that the former supremum is finite as well. From the dominated convergence theorem, we then have

\[
\sum_{\sigma'} \lim_{t \to 0} \frac{1}{t} \left[ \int_{s'} \left[ \Theta^{(i)}(x_\sigma(s + t), x_{\sigma'}(s' + t)) - \Theta^{(i)}(x_\sigma(s), x_{\sigma'}(s')) \right] h(x_{\sigma'}(s' + t)) ds'
\]

\[
= \sum_{\sigma'} \int_{s'} \lim_{t \to 0} \frac{1}{t} \left[ \Theta^{(i)}(x_\sigma(s + t), x_{\sigma'}(s' + t)) - \Theta^{(i)}(x_\sigma(s), x_{\sigma'}(s')) \right] h(x_{\sigma'}(s' + t)) ds'
\]

\[
= \sum_{\sigma'} \int_{s'} \Theta^{(i+1)}(x_\sigma(s), x_{\sigma'}(s')) h(x_{\sigma'}(s')) ds'
\]

(E.69)
Similarly, as \( \text{ess sup}_{x \in \mathcal{M}} \left| \frac{d}{ds} h(x) \right| \) is finite and \( \mathcal{M} \) is compact, from the dominated convergence theorem we also have

\[
\sum_{\sigma} \lim_{t \to 0} \frac{1}{t} \int_{s'} \Theta^{(i)}(x_{\sigma}(s), x_{\sigma}(s')) [h(x_{\sigma'}(s') + t) - h(x_{\sigma'}(s'))] \, ds' = \sum_{\sigma} \int \lim_{t \to 0} \frac{1}{t} \Theta^{(i)}(x_{\sigma}(s), x_{\sigma}(s')) [h(x_{\sigma'}(s') + t) - h(x_{\sigma'}(s'))] \, ds' \\
= \sum_{\sigma} \int \Theta^{(i)}(x_{\sigma}(s), x_{\sigma}(s')) \frac{d}{ds} h(x_{\sigma'}(s')) \, ds'
\]

(E.70)

Summing (E.69) and (E.70) shows the claim. ■

**Lemma E.22.** There is an absolute constant \( C > 0 \) such that for all \( L \geq C \), any differentiable \( h : \mathcal{M} \to \mathbb{C} \) and any \( \varepsilon \in (0, \frac{1}{2}] \), if the operator \( \widehat{M}_L \) and the subspace \( S_\varepsilon \) are as defined in (E.11) and (E.12), then we have

\[
\frac{d}{ds} P_{S_\varepsilon} [h] = P_{S_\varepsilon} \left[ \frac{d}{ds} h \right], \\
\frac{d}{ds} \widehat{M}_L P_{S_\varepsilon} [h] = \widehat{M}_L P_{S_\varepsilon} \left[ \frac{d}{ds} h \right].
\]

Also, suppose the hypotheses of Lemma E.6 are satisfied, so that \( P_{S_\varepsilon} \widehat{M}_L P_{S_\varepsilon} \) is invertible over \( S_\varepsilon \). Then one has in particular

\[
\frac{d}{ds} \left( P_{S_\varepsilon} \widehat{M}_L P_{S_\varepsilon} \right)^{-1} [h] = \left( P_{S_\varepsilon} \widehat{M}_L P_{S_\varepsilon} \right)^{-1} \left[ \frac{d}{ds} h \right].
\]

**Proof.** The condition on \( L \) implies that \( \widehat{M}_L \) is well-defined. For any operator \( T \) that diagonalizes in the Fourier basis for \( S_\varepsilon \), i.e. for any \( h \in L^2(\mathcal{M}) \), \( T \) satisfies

\[
T[h] = \sum_{\sigma \in \{\pm\}} \sum_{k = -K_{\varepsilon,\sigma}}^{K_{\varepsilon,\sigma}} m_{\sigma,k} \phi_{\sigma,k} \phi_{\sigma,k}^* h
\]

(E.71)

for some coefficients \( m_{\sigma,k} \in \mathbb{C} \) independent of \( h \),\(^7\) we have

\[
\frac{d}{ds} T[h] = \sum_{\sigma \in \{\pm\}} \sum_{k = -K_{\varepsilon,\sigma}}^{K_{\varepsilon,\sigma}} m_{\sigma,k} \left( \frac{d}{ds} \phi_{\sigma,k} \right) \phi_{\sigma,k}^* h \\
= \sum_{\sigma \in \{\pm\}} \sum_{k = -K_{\varepsilon,\sigma}}^{K_{\varepsilon,\sigma}} m_{\sigma,k} \frac{i2\pi k}{\text{len}(\mathcal{M}_\sigma)} \phi_{\sigma,k} \phi_{\sigma,k}^* h,
\]

where we recall the definition of the Fourier basis functions from (C.11) for the second equality. Now fix \( h \) differentiable as in the statement of the lemma. On the other hand, since \( \phi_{\sigma,k}^* h \) is simply some complex number, which does not depend on \( s \), we have

\[
\left( \frac{d}{ds} \phi_{\sigma,k} \right)^* h + \phi_{\sigma,k}^* \frac{d}{ds} h = 0,
\]

and so

\[
T \left[ \frac{d}{ds} h \right] = \sum_{\sigma \in \{\pm\}} \sum_{k = -K_{\varepsilon,\sigma}}^{K_{\varepsilon,\sigma}} m_{\sigma,k} \phi_{\sigma,k} \phi_{\sigma,k}^* \frac{d}{ds} h
\]

\(^7\) Here and in the sequel, we recall that we are using the notation \( \phi_{\sigma,k}^* h = \langle \phi_{\sigma,k}^*, h \rangle = \int_{\mathcal{M}} \overline{\phi_{\sigma,k}(x)} h(x) \, dx \) for the standard inner product on complex-valued functions on \( \mathcal{M} \).
We have for any differentiable function \( f : \mathcal{M} \to \mathbb{C} \)
\[
\left\| \frac{d}{ds} P_{S_\varepsilon} f \right\|_{L^2} \leq \frac{\sqrt{c}}{r_\varepsilon} \left\| P_{S_\varepsilon} f \right\|_{L^2},
\]
\[
\left\| P_{S_\varepsilon} f \right\|_{L^2} \leq \frac{2r_\varepsilon}{\sqrt{c}} \left\| \frac{d}{ds} P_{S_\varepsilon} f \right\|_{L^2}.
\]

**Proof.** The condition on \( L \) guarantees that \( \hat{M}_\varepsilon \) is well-defined. From (E.12), \( S_\varepsilon = S_{K_\varepsilon,+,K_\varepsilon,-} \) with \( K_{\varepsilon,\sigma} = \left[ \frac{\varepsilon^{1/2} \min \{ \text{len}(M_+), \text{len}(M_-) \}}{2\pi r_\varepsilon} \right] \) for \( \sigma \in \{ +, - \} \), then by orthonormality of the Fourier basis functions (C.11), we have
\[
\left\| \frac{d}{ds} P_{S_\varepsilon} f \right\|_{L^2} = \left\| \frac{d}{ds} \sum_{\sigma=\pm} \sum_{k=-K_{\varepsilon,\sigma}}^{K_{\varepsilon,\sigma}} \phi_{\sigma,k} \phi_{\sigma,k}^* f \right\|_{L^2}
\]
\[
= \left\| \sum_{\sigma=\pm} \sum_{k=-K_{\varepsilon,\sigma}}^{K_{\varepsilon,\sigma}} \frac{i 2\pi k}{\text{len}(M_\sigma)} \phi_{\sigma,k} \phi_{\sigma,k}^* f \right\|_{L^2}
\]
\[
\leq \frac{\sqrt{c}}{r_\varepsilon} \left( \sum_{\sigma=\pm} \sum_{k=-K_{\varepsilon,\sigma}}^{K_{\varepsilon,\sigma}} \| \phi_{\sigma,k} \phi_{\sigma,k}^* f \|^2_{L^2} \right)^{1/2}
\]
\[
= \frac{\sqrt{c}}{r_\varepsilon} \left\| P_{S_\varepsilon} f \right\|_{L^2}.
\]

Above, the inequality follows because \( |k| \leq K_{\varepsilon,\sigma} \) implies \( 2\pi |k| / \text{len}(M_\sigma) \leq \sqrt{c}/r_\varepsilon \), and because the Fourier basis functions are mutually orthogonal (and \( \| f \|^2_{L^2} = \langle f, f \rangle \)). This establishes the first claim.

For the second claim, we have
\[
P_{S_\varepsilon^\pm} f = \sum_{\sigma=\pm} \sum_{k \geq K_{\varepsilon,\sigma}} \phi_{\sigma,k} \phi_{\sigma,k}^* f.
\]

When \( L \geq \left[ \frac{\varepsilon^{1/2} \min \{ \text{len}(M_+), \text{len}(M_-) \}}{12\pi^2 L^{4\varepsilon+1}} \right] \geq 1 \) and thus \( K_{\varepsilon,\pm} = \left[ \frac{\varepsilon^{1/2} \min \{ \text{len}(M_+), \text{len}(M_-) \}}{2\pi r_\varepsilon} \right] \geq \left[ \frac{\varepsilon^{1/2} \min \{ \text{len}(M_+), \text{len}(M_-) \}}{4\pi r_\varepsilon} \right] \), whence
\[
\left\| \frac{d}{ds} P_{S_\varepsilon^\pm} f \right\|_{L^2} = \left\| \sum_{\sigma=\pm} \sum_{k \geq K_{\varepsilon,\sigma}} \left( \frac{d}{ds} \phi_{\sigma,k} \right) \phi_{\sigma,k}^* f \right\|_{L^2}
\]
Furthermore, for \( \varepsilon \) in force, for any \( \zeta \), where the difference quotient involving \( \phi_{\sigma,k} \) is bounded by \( O(\varepsilon) \), which together with the extra smoothness of \( f \) leads to an integrable upper bound.

**Definition E.24.** For any \( \varepsilon \in (0, \frac{3}{4}) \), let \( P_1 = P_1(M_4, \text{len}(\mathcal{M}), \Delta_{\varepsilon}^{-1}) \), \( P_2 = P_2(M_4, M_5, \text{len}(\mathcal{M}), \Delta_{\varepsilon}^{-1}) \), \( P_3 = P_3(M_3, M_4, M_5, \text{len}(\mathcal{M}), \Delta_{\varepsilon}^{-1}) \) as defined in Lemmas E.18 to E.20. We let \( \Phi(C_\zeta, \varepsilon) \) for some constant \( C_\zeta > 0 \) denote the set of all functions \( \zeta \in C^3(\mathcal{M}) \) which satisfy

\[
\| \zeta \|_{L^2} \leq C_\zeta \\
\| \zeta^{(1)} \|_{L^2} \leq C_\zeta \frac{P_1}{\log L} \\
\| \zeta^{(2)} \|_{L^2} \leq C_\zeta \left( \frac{P_2}{\log L} + \frac{P_2^2}{\log^2 L} \right) \\
\| \zeta^{(3)} \|_{L^2} \leq C_\zeta \left( \frac{P_3}{\log L} + \frac{P_2 P_3}{\log^2 L} + \frac{P_3^3}{\log^3 L} \right).
\]

Furthermore, for \( \varepsilon \geq \varepsilon' > 0 \), one has \( \Delta_{\varepsilon}^{-1} \leq \Delta_{\varepsilon'}^{-1} \). As \( P_1, P_2, P_3 \) have positive coefficients, \( \zeta \in \Phi(C_\zeta, \varepsilon) \) implies \( \zeta \in \Phi(C_\zeta, \varepsilon') \), i.e., \( \Phi(C_\zeta, \varepsilon') \subseteq \Phi(C_\zeta, \varepsilon) \).

**Lemma E.25.** For any \( \varepsilon \in (0, \frac{3}{4}) \), there exist numbers \( C_\zeta, C'_\zeta > 0 \) such that when the conditions of Theorem E.1 are in force, for any \( \zeta \in S_z \cap \Phi(C_\zeta, \varepsilon) \), the certificate \( g_\varepsilon[\zeta] \) defined in Theorem E.1 satisfies

\[
g_\varepsilon[\zeta] \in \Phi \left( \frac{C_\zeta C'_\zeta}{n \log L}, \varepsilon \right). \tag{E.72}
\]

**Proof.** Following Lemma E.22 and Lemma E.21, we have that

\[
g_\varepsilon^{(1)}[\zeta] = \sum_{\ell = 1}^{\infty} \left( \sum_{a=0}^{\ell-1} \left( \begin{array}{l}
(P_{S_z} \tilde{M}_z P_{S_z})^{-1} P_{S_z} \Theta^o \tilde{M}_z P_{S_z} \end{array} \right)^a \left( \begin{array}{l}
P_{S_z} \tilde{M}_z P_{S_z} \end{array} \right)^{-1} P_{S_z} \Theta^{(1)} P_{S_z} \\
\times \left( \begin{array}{l}
P_{S_z} \tilde{M}_z P_{S_z} \end{array} \right)^{-1} P_{S_z} \Theta^o \tilde{M}_z P_{S_z} \end{array} \right)^{\ell-a-1} \right) + g_\varepsilon[\zeta^{(1)}].
\]

As \( \zeta \in S_z \), we have \( \zeta^{(1)} = \frac{d}{ds} \zeta = \frac{d}{ds} P_{S_z} \zeta = P_{S_z} \frac{d}{ds} \zeta \in S_z \), and thus following Theorem E.1, there exists constant \( c \) such that

\[
\| g_\varepsilon[\zeta^{(1)}] \|_{L^2} \leq \frac{\| \zeta^{(1)} \|_{L^2}}{\varepsilon c n \log L}.
\]
From Lemma E.6, there exists $C, c' > 0$ such that when $L \geq C$,

$$
\lambda_{\min}(P_S \tilde{M}_s P_{S_s}) \geq c n \log L.
$$

Under the conditions of Theorem E.1, we have

$$
\left\| \left( P_{S_s} \tilde{M}_s P_{S_s} \right)^{-1} P_{S_s} (\Theta^o - \tilde{M}_s) P_{S_s} \right\|_{L^2 \to L^2} \leq 1 - \varepsilon.
$$

Let $P_1 = P_1 \left( M_4, \text{len}(M), \Delta^{-1}_\varepsilon \right)$, $P_2 = P_2 \left( M_4, M_5, \text{len}(M), \Delta^{-1}_\varepsilon \right)$, and $P_3 = P_3 \left( M_3, M_4, M_5, \text{len}(M), \Delta^{-1}_\varepsilon \right)$ be the polynomials in Lemmas E.18 to E.20. From Lemma E.18, we have

$$
\| g^{(1)}_\varepsilon [\zeta] \|_{L^2} \leq \sum_{\ell=1}^{\ell-1} \sum_{a=0}^{a} (1 - \varepsilon)^{\ell-1} \| \Theta^{(1)} \|_{L^2 \to L^2} \| \zeta \|_{L^2} \left( \frac{\varepsilon n \log L}{(c n \log L)^2} \right)
$$

$$
+ \sum_{\ell=0}^{\ell} \sum_{a=0}^{a} (1 - \varepsilon)^{\ell-1} \| \zeta^{(1)} \|_{L^2} \left( \frac{\varepsilon n \log L}{(c n \log L)^2} \right)
$$

$$
= \sum_{\ell=0}^{\ell} \sum_{a=0}^{a} (1 - \varepsilon)^{\ell-1} \| \zeta^{(1)} \|_{L^2} \left( \frac{\varepsilon n \log L}{(c n \log L)^2} \right)
$$

$$
\leq \frac{P_1 \| \zeta \|_{L^2}}{\varepsilon^2 c^2 n \log^2 L} + \frac{\| \zeta^{(1)} \|_{L^2}}{\varepsilon n \log L}.
$$

(E.73)

From the fact that $\zeta \in \Phi (C_\varepsilon, \varepsilon)$, we further obtain

$$
\| g^{(1)}_\varepsilon [\zeta] \|_{L^2} \leq \frac{P_1 C_\varepsilon}{\varepsilon^2 c^2 n \log^2 L} + \frac{C_\varepsilon}{\varepsilon n \log L} \log L
$$

$$
\leq C_\varepsilon \frac{P_1 C_\varepsilon}{n \log^2 L}.
$$

For the second derivative, we have

$$
g^{(2)}_\varepsilon [\zeta] = \sum_{\ell=1}^{\ell} \sum_{a=0}^{a} \left( \left( P_S \tilde{M}_s P_{S_s} \right)^{-1} P_{S_s} (\Theta^o - \tilde{M}_s) P_{S_s} \right)^a
$$

$$
\times \left( P_S \tilde{M}_s P_{S_s} \right)^{-1} P_{S_s} (\Theta^{(2)} P_{S_s})
$$

$$
\times \left( \left( P_S \tilde{M}_s P_{S_s} \right)^{-1} P_{S_s} (\Theta^o - \tilde{M}_s) P_{S_s} \right)^{\ell-a-1} \left( P_S \tilde{M}_s P_{S_s} \right)^{-1} \zeta
$$

$$
+ 2 \sum_{\ell=1}^{\ell} \sum_{a=0}^{a} \sum_{a' = a}^{a} \left( \left( P_S \tilde{M}_s P_{S_s} \right)^{-1} P_{S_s} (\Theta^o - \tilde{M}_s) P_{S_s} \right)^a
$$

$$
\times \left( P_S \tilde{M}_s P_{S_s} \right)^{-1} P_{S_s} \Theta^{(1)} P_{S_s}
$$

$$
\times \left( \left( P_S \tilde{M}_s P_{S_s} \right)^{-1} P_{S_s} (\Theta^o - \tilde{M}_s) P_{S_s} \right)^{a'-a}
$$

$$
\times \left( P_S \tilde{M}_s P_{S_s} \right)^{-1} P_{S_s} \Theta^{(1)} P_{S_s}
$$

$$
\times \left( \left( P_S \tilde{M}_s P_{S_s} \right)^{-1} P_{S_s} (\Theta^o - \tilde{M}_s) P_{S_s} \right)^{\ell-a'-2} \left( P_S \tilde{M}_s P_{S_s} \right)^{-1} \zeta
$$

$$
+ 2 g^{(1)}_\varepsilon [\zeta^{(1)}] - g^{(2)}_\varepsilon [\zeta] + g^{(2)}_\varepsilon [\zeta].
$$

46
From (E.73), as $\zeta^{(1)}, \zeta^{(2)} \in S$, we have
\[
\|g_\varepsilon^{(1)}[\zeta^{(1)}]\|_{L^2} \leq \frac{P_1 \|\zeta^{(1)}\|_{L^2}}{\varepsilon^2 c^2 n \log^2 L} + \frac{\|\zeta^{(2)}\|_{L^2}}{\varepsilon c n \log L},
\]
\[
\|g_\varepsilon[\zeta^{(2)}]\|_{L^2} \leq \frac{\|\zeta^{(2)}\|_{L^2}}{\varepsilon c n \log L}.
\]
which leads to
\[
\|g_\varepsilon^{(2)}[\zeta]\|_{L^2} \leq \frac{\|\Theta^{(2)}\|_{L^2 \to L^2} \|\zeta\|_{L^2}}{\varepsilon^2 (c n \log L)^2} + \sum_{\ell=1}^{\infty} \ell (\ell - 1) (1 - \varepsilon) \ell - 2 \frac{\|\Theta^{(1)}\|_{L^2 \to L^2} \|\zeta\|_{L^2}}{(c n \log L)^3}
\]
\[
+ 2 \|g_\varepsilon^{(1)}[\zeta^{(1)}]\|_{L^2} + \|g_\varepsilon[\zeta^{(2)}]\|_{L^2}
\]
\[
\leq \frac{\|\Theta^{(2)}\|_{L^2 \to L^2} \|\zeta\|_{L^2}}{\varepsilon^2 (c n \log L)^2} + \frac{2\|\Theta^{(1)}\|_{L^2 \to L^2} \|\zeta\|_{L^2}}{\varepsilon^3 (c n \log L)^3}
\]
\[
+ 2 \left( \frac{P_1 \|\zeta^{(1)}\|_{L^2}}{\varepsilon^2 c^2 n \log^2 L} + \frac{\|\zeta^{(2)}\|_{L^2}}{\varepsilon c n \log L} \right) + \frac{\|\zeta^{(2)}\|_{L^2}}{\varepsilon c n \log L}
\]
\[
\leq \frac{P_2 \|\zeta\|_{L^2}}{\varepsilon^2 c^2 n \log^2 L} + \frac{2P_2 \|\zeta\|_{L^2}}{\varepsilon^3 c^3 n \log^3 L} + \frac{2P_1 \|\zeta^{(1)}\|_{L^2}}{\varepsilon^2 c n \log^2 L} + \frac{3\|\zeta^{(2)}\|_{L^2}}{\varepsilon c n \log L}.
\]
(E.74)

Again as $\zeta \in \Phi(C_\zeta, \varepsilon)$ we have
\[
\|g_\varepsilon^{(2)}[\zeta]\|_{L^2} \leq \frac{P_2 C_\zeta}{\varepsilon^2 c^2 n \log^2 L} + \frac{2P_2^2 C_\zeta}{\varepsilon^3 c^3 n \log^3 L}
\]
\[
+ \frac{2P_1 C_\zeta}{\varepsilon^2 c^2 n \log^2 L} + \frac{C_\zeta}{\varepsilon c n \log L} \left( \frac{P_2}{\log L} + \frac{P_2^2}{\log^2 L} \right)
\]
\[
\leq C_\zeta C_\varepsilon \left( \frac{P_2}{n \log^3 L} + \frac{P_2}{n \log^2 L} \right).
\]

For third derivative, we have
\[
g_\varepsilon^{(3)}[\zeta] = \sum_{\ell=1}^{\infty} (-1)^{\ell - 1} \sum_{a=0}^{\ell - 1} \left( P_S \hat{M}_a P_{S^a} \right)^{-1} P_{S^a} (\Theta^\circ - \hat{M}) P_{S^a} \]
\[
\times \left( P_S \hat{M}_a P_S \right)^{-1} P_S (\Theta^\circ - \hat{M}) P_S
\]
\[
\times \left( P_S \hat{M}_a P_{S^a} \right)^{-1} P_{S^a} (\Theta^\circ - \hat{M}) P_{S^a}
\]
\[
+ 3 \sum_{\ell=1}^{\infty} (-1)^{\ell - 2} \sum_{a=0}^{\ell - 2} \sum_{a' = a} \left( P_S \hat{M}_a P_{S^a} \right)^{-1} P_{S^a} (\Theta^\circ - \hat{M}) P_{S^a}
\]
\[
\times \left( P_S \hat{M}_a P_{S^a} \right)^{-1} P_{S^a} (\Theta^\circ - \hat{M}) P_{S^a}
\]
\[
\times \left( P_S \hat{M}_a P_{S^a} \right)^{-1} P_{S^a} (\Theta^\circ - \hat{M}) P_{S^a}
\]
\[
+ 3 \sum_{\ell=1}^{\infty} (-1)^{\ell - 2} \sum_{a=0}^{\ell - 2} \sum_{a' = a} \left( P_S \hat{M}_a P_{S^a} \right)^{-1} P_{S^a} (\Theta^\circ - \hat{M}) P_{S^a}
\]
\[
\times \left( P_S \hat{M}_a P_{S^a} \right)^{-1} P_{S^a} (\Theta^\circ - \hat{M}) P_{S^a}
\]

47
Similarly, as \( \zeta^{(1)}, \zeta^{(2)} \) and \( \zeta^{(3)} \in S_c \), plug in results in (E.73) and (E.74) we can control

\[
\| g_{\varepsilon}^{(3)}[\zeta] \|_{L^2} \leq \frac{\| \Theta^{(3)} \|_{L^2 \to L^2} \| \zeta \|_{L^2}}{\varepsilon^2 (cn \log L)^2} + 3 \| \Theta^{(2)} \|_{L^2 \to L^2} \| \Theta^{(1)} \|_{L^2 \to L^2} \| \zeta \|_{L^2} \varepsilon^3 (cn \log L)^3 \\
+ \frac{3 \| \Theta^{(2)} \|_{L^2 \to L^2} \| \Theta^{(1)} \|_{L^2 \to L^2} \| \zeta \|_{L^2}}{\varepsilon^2 (cn \log L)^2} + \frac{3 \| \Theta^{(2)} \|_{L^2 \to L^2} \| \Theta^{(1)} \|_{L^2 \to L^2} \| \zeta \|_{L^2}}{\varepsilon^3 (cn \log L)^3} \\
+ \frac{3 \| \Theta^{(3)} \|_{L^2 \to L^2} \| \zeta \|_{L^2}}{\varepsilon^2 (cn \log L)^2} + \frac{3 \| \Theta^{(2)} \|_{L^2 \to L^2} \| \Theta^{(1)} \|_{L^2 \to L^2} \| \zeta \|_{L^2}}{\varepsilon^3 (cn \log L)^3} \\
+ \frac{6 \| \Theta^{(1)} \|_{L^2 \to L^2} \| \zeta \|_{L^2}}{\varepsilon^3 (cn \log L)^3} \\
+ \frac{\| \zeta \|_{L^2}}{\varepsilon^2 c^2 n \log^2 L} \left( \frac{6 P_{3}^2}{\log L} + 3 P_{2} \right) + \frac{9 P_{1} \| \zeta \|_{L^2}}{\varepsilon^2 c^2 n \log^2 L} + \frac{13 \| \zeta \|_{L^2}}{\varepsilon^2 c^2 n \log^2 L} \\
\leq \frac{P_{3} \| \zeta \|_{L^2}}{\varepsilon^2 c^2 n \log^2 L} + \frac{3 P_{3} P_{1} \| \zeta \|_{L^2}}{\varepsilon^3 c^3 n \log^3 L} + \frac{9 P_{1} \| \zeta \|_{L^2}}{\varepsilon^2 c^2 n \log^2 L} + \frac{13 \| \zeta \|_{L^2}}{\varepsilon^2 c^2 n \log^2 L}
\]
Plug in bounds for norms of $\zeta^{(1)}$, $\zeta^{(2)}$ and $\zeta^{(3)}$ we get
\[
\|g^{(3)}(\zeta)\|_{L^2} \leq C_\epsilon C_\zeta \left( \frac{P_1^3}{n \log^3 L} + \frac{P_1 P_2}{n \log^4 L} + \frac{P_3}{n \log^2 L} \right).
\]
Combined with zero’s order condition of $g, [\zeta]$, which follows directly from Theorem E.1, and we know that there exists $C_\epsilon$ such that $g \in \Phi\left(\frac{C_\zeta C_\zeta}{n \log L}, \epsilon\right)$. ■

**Lemma E.26.** For $\epsilon \in (0, \frac{3}{4})$, when $L$ satisfies conditions in Theorem E.1, there exists positive constant $C$ such that
\[
\|\Theta^o\|_{L^2 \to L^2} \leq C n \log(L).
\]

**Proof.** As $\hat{M}_\epsilon$ in (E.11) is invariant in Fourier basis as shown in Lemma E.6, we have
\[
\|\hat{P}_S \hat{M}_\epsilon P_{S_\epsilon}\|_{L^2 \to L^2} \leq \|\hat{M}_\epsilon\|_{L^2 \to L^2}.
\]
From Lemma E.2,
\[
\|\Theta^o\|_{L^2 \to L^2} = \|\hat{M}_\epsilon + \Theta^o - \hat{M}_\epsilon\|_{L^2 \to L^2}
\]
\[
\leq \|\hat{M}_\epsilon\|_{L^2 \to L^2} + \|\Theta^o - \hat{M}\|_{L^2 \to L^2}
\]
\[
\leq \|\hat{M}_\epsilon\|_{L^2 \to L^2} + (1 - \epsilon) \lambda_{\min}(\hat{P}_S \hat{M}_\epsilon P_{S_\epsilon})
\]
\[
\leq 2 \|\hat{M}_\epsilon\|_{L^2 \to L^2}.
\]
As when $L \geq (\epsilon^{-1/2} (6 \pi k)^{\frac{1}{2}})^{-\frac{1}{2}} n^2 a_c r_c$ we have $r_c \leq \sqrt{\frac{\pi}{k}}$, where $a_c, r_c$ are defined in (E.7) and (E.8), and following Lemma E.4 we have
\[
\angle(x_\sigma(s), x_\sigma(s')) \geq (1 + \epsilon) |s - s'|
\]
for any $x_\sigma(s)$ and $|s - s'| \leq r_c$. Then follow Lemma E.5 and (F.11) Lemma F.9 and monotonicity of $\psi^o$ in Lemma G.5, we have
\[
\|\hat{M}_\epsilon\|_{L^2 \to L^2} \leq \max_{x_\sigma(s) \in \mathcal{M}} \int_{s' = s - r_c}^{s + r_c} \psi^o(\angle x_\sigma(s), x_\sigma(s')) ds'
\]
\[
\leq \max_{x_\sigma(s) \in \mathcal{M}} \int_{s' = s - r_c}^{s + r_c} \psi^o((1 - \epsilon) |s' - s|) ds'
\]
\[
= (1 - \epsilon)^{-1} \int_{t = -(1 - \epsilon) r_c}^{2} \psi^o(t) dt
\]
\[
\leq (1 - \epsilon)^{-1} C \log \left( 1 + \frac{(L - 3)(1 - \epsilon)r_c}{3\pi} \right)
\]
\[
\leq C' n \log L
\]
for some constant $C$, which concludes the claim. ■

**Lemma E.27.** For any $\epsilon \in (0, \frac{3}{4})$, there exists constant $C \geq 0$ such that for any $g \in \Phi(C_\epsilon, \epsilon)$, under the conditions of Theorem E.1, we have
\[
\Theta^o[g] \in \Phi(C_\epsilon C_\zeta n \log L, \epsilon)
\]
As a consequence, for $\zeta \in S_\epsilon \cap \Phi(C_\zeta, \epsilon)$ letting $g_\zeta[\zeta]$ be the certificate in the statement of Theorem E.1, there exists number $C_\epsilon \geq 0$ such that
\[
\Theta^o[g_\zeta]\zeta) - \zeta \in \Phi(C_\epsilon C_\zeta, \epsilon).
\]

**Proof.** From Lemma E.26 we have $\|\Theta^o\|_{L^2 \to L^2} \leq C n \log(L)$ for some constant $C$. Let $P_1, P_2, P_3$ be the polynomials in Lemmas E.18 to E.20. Following Lemmas E.18 to E.21 and the fact that $g \in \Phi(C_\epsilon, \epsilon)$, we have following control for derivatives of $\Theta^o[g]$:
\[
\left\| \frac{d}{ds} \Theta^o[g] \right\|_{L^2 \to L^2} = \left\| \Theta^o[g^{(1)}] + \Theta^{(1)}[g] \right\|_{L^2 \to L^2}
\]
\[ \left\| \frac{d^2}{ds^2} \Theta^s[g] \right\|_{L^2 \to L^2} \leq Cn \log L \left( \frac{P_1 C_2}{\log L} + 2 \frac{P_2 C_2}{\log^2 L} \right) + P_1 n \frac{P_1 C_2}{\log L} + P_2 n C_g = (C + 1) P_1 n C_g, \]

\[ \left\| \frac{d^3}{ds^3} \Theta^s[g] \right\|_{L^2 \to L^2} \leq Cn \log L \left( \frac{P_3 C_2}{\log L} + 3 P_1 n \left( \frac{P_2 C_2}{\log L} + \frac{P_2^2 C_2}{\log^2 L} \right) \right) + 3 P_2 n \frac{P_1 C_2}{\log L} + P_3 n C_g = (C + 3) \left( \frac{P_3 C_2}{\log L} + \frac{P_1 P_2 C_2}{\log L} + \frac{P_3^3 C_2}{\log^3 L} \right) n. \]

which leads to \( \Theta^s[g] \in \Phi((C + 3) C_g n \log L, \varepsilon) \) and finish the claim. The other part of the claim follows directly from Lemma E.25 as \( g_e[\zeta] \in \Phi \left( \frac{C_e C_p}{n \log L}, \varepsilon \right) \).

**Lemma E.28.** Let \( \varepsilon \in (0, \frac{3}{4}) \), \( a_\varepsilon, r_\varepsilon, S_\varepsilon \) be as in (E.7), (E.8) and (E.12). There exist numbers \( C_\varepsilon, C'_\varepsilon \) such that when \( L \geq \left( \varepsilon^{1/2} \min\{\text{len}(\mathcal{M}_+), \text{len}(\mathcal{M}_-)\}/(12 \pi^2) \right)^{-\frac{a_\varepsilon + 1}{4}} \) and the conditions of Theorem E.1 are in force, for any \( w \in \Phi(C_w, \varepsilon) \), we have

\[ \| P_{S_{\varepsilon}^w} w \|_{L^2} \leq C_{\varepsilon} C_w \frac{r_{\varepsilon}^3}{\log L} \left( P_3 + \frac{P_1 P_2}{\log L} + \frac{P_3^3}{\log^3 L} \right). \]

where \( P_1, P_2, P_3 \) are the polynomials from Lemmas E.18 to E.20 respectively.

As a consequence, for \( \zeta \in S_\varepsilon \cap \Phi(C_\varepsilon, \varepsilon) \), letting \( g_e[\zeta] \) be as in Theorem E.1, we have

\[ \| \Theta[g_e[\zeta]] - \zeta \|_{L^2} \leq C_\varepsilon' C_\zeta \frac{r_{\varepsilon}^3}{\log L} \left( P_3 + \frac{P_1 P_2}{\log L} + \frac{P_3^3}{\log^3 L} \right). \]

for some \( C_\varepsilon' > 0 \).

**Proof.** When \( L \geq \left( \varepsilon^{1/2} \min\{\text{len}(\mathcal{M}_+), \text{len}(\mathcal{M}_-)\}/(12 \pi^2) \right)^{-\frac{a_\varepsilon + 1}{4}} \), from Lemma E.23 we have

\[ \| P_{S_{\varepsilon}^w} w \| \leq \left( \frac{2r_{\varepsilon}}{\sqrt{\varepsilon}} \right)^3 \| P_{S_{\varepsilon}^w} w^{(3)} \|_{L^2} \]

\[ \leq \left( \frac{2r_{\varepsilon}}{\sqrt{\varepsilon}} \right)^3 \| w^{(3)} \|_{L^2} \]

\[ \leq C_w \left( \frac{2r_{\varepsilon}}{\sqrt{\varepsilon}} \right)^3 \left( \frac{P_3}{\log L} + \frac{P_2 P_1}{\log^2 L} + \frac{P_3^3}{\log^3 L} \right). \]

As \( \zeta \in \Phi(C_\varepsilon, \varepsilon) \), from Lemma E.27 we get \( \Theta^s[g] - \zeta \in \Phi(C'_\varepsilon C_\zeta, \varepsilon) \) for some \( C'_\varepsilon > 0 \). The rest follows from the fact that \( P_{S_{\varepsilon}} (\Theta^s[g] - \zeta) = 0 \) and thus \( \| \Theta[g_e[\zeta]] - \zeta \|_{L^2} = \| P_{S_{\varepsilon}} (\Theta^s[g] - \zeta) \|_{L^2} \).

\[ \| \Theta^s[g] - \zeta \|_{L^2} = \| P_{S_{\varepsilon}} (\Theta^s[g] - \zeta) \|_{L^2}. \]
E.3 Certificates with Density and DC

Proof Sketch and Organization. In this section, we leverage the calculations in the previous sections to prove Theorem D.2, which gives a near solution to the equation

\[ \Theta_\mu[g](x) = \int \Theta(x, x') g(x') \rho(x') dx' = \zeta(x). \]

To accomplish this, we need to account for two factors: the presence of a constant (DC) term in \( \Theta(x, x') = \Theta^0(x, x') + \psi(\pi), \) and the presence of the data density \( \rho \) in \( \Theta_\mu. \)

Our approach is conceptually straightforward: since \( \Theta = \Theta^0 + \psi(\pi) \mathbb{1} \mathbb{1}^*, \) we produce near solutions to two equations

\[ \Theta^0[g](x) = \zeta(x), \]
\[ \Theta^0[g_1](x) = 1, \]

and then combine them to nearly solve \( \Theta[h] = \zeta, \) by setting \( h = g + \alpha g_1 \) for an appropriate choice of \( \alpha, \)

\[ \alpha = -\frac{\psi(\pi) \mathbb{1}[g]}{\psi(\pi) \mathbb{1}[g_1] + 1}. \] (E.75)

Here and in the rest of this section, we write \( \mathbb{1}[g] \) to denote \( \mathbb{1}^* g. \)

The statement of Theorem D.2 makes two demands on \( h: \) small approximation error \( \|\Theta[h] - \zeta\|_{L^2} \) and small size \( \|h\|_{L^2}. \) These demands introduce a tension, which forces us to work with DC subtracted solutions \( g_l[\zeta] \) defined in (E.13) at multiple scales \( \varepsilon. \) We will set \( g = g_{\varepsilon_1}[\zeta] \) with \( \varepsilon_0 \) small, which ensures that both \( \|\Theta^0[g] - \zeta\|_{L^2} \) and \( \|g\|_{L^2} \) are small. We would like to similarly set \( g_1 = g_{\varepsilon_1}[\zeta_1], \) with \( \zeta_1 \equiv 1. \) In order to ensure that \( h \) is small, we need to ensure that the coefficient \( \alpha \) defined in (E.75) is also small, which in turn requires a lower bound on \( \mathbb{1}[g_1]. \) This is straightforward if \( g_1 \) is (pointwise) nonnegative, but challenging if \( g_1 \) can take on arbitrary signs. The function \( g_1 \) is defined by the Neumann series

\[ g_1 = g_{\varepsilon_1}[\zeta_1] \]
\[ = \sum_{\ell=0}^{\infty} (-1)^\ell \left( P_{S_{\varepsilon_1}} \widehat{M}_{\varepsilon_1} P_{S_{\varepsilon_1}} \right)^{-1} P_{S_{\varepsilon_1}} \left( \Theta^0 - \widehat{M}_{\varepsilon_1} \right) P_{S_{\varepsilon_1}} \] \[ \left( P_{S_{\varepsilon_1}} \widehat{M}_{\varepsilon_1} P_{S_{\varepsilon_1}} \right)^{-1} \zeta_1. \]

Although this expression is complicated, the first (\( \ell = 0 \)) summand is always nonnegative. If we choose \( \varepsilon_1 \) large, this expression will be dominated by the first term, providing the necessary control on \( \mathbb{1}[g_1]. \) So, we will use two different scales, \( \varepsilon_0 < \varepsilon_1 \) in constructing \( g \) and \( g_1, \) respectively.

The issue introduced by the use of a large scale \( \varepsilon_1 \) is that the approximation error \( \|\Theta^0[g_1] - \zeta_1\|_{L^2} \) is not sufficiently small for our purposes. To address this issue, we introduce an iterative construction, which produces a sequence of increasingly accurate solutions \( h_{(\varepsilon)}, \) each of which removes some portion of the approximation error in the previous solution. This sequence converges to our promised certificate \( h. \)

More concretely, we will set \( \varepsilon_0 = \frac{1}{25} \) and \( \varepsilon_1 = \frac{1}{100}. \) For parameters \( a, r \) defined in (E.7) and (E.8), these choices of \( \varepsilon \) ensure that \( a_{\varepsilon_0} > \frac{4}{5} \) and \( a_{\varepsilon_1} > \frac{1}{5}, \) and so

\[ r_{\varepsilon_0} < 6\pi L^{-\frac{4}{5}}, \]
\[ r_{\varepsilon_1} < 6\pi L^{-\frac{1}{10}}. \]

We further choose \( \delta_0 = 1 - \varepsilon_0 \) and \( \delta_1 = \delta_0 \sqrt{\varepsilon_0}/\sqrt{\varepsilon_1} < 1 - \varepsilon_1. \) This setting satisfies \( \delta_0 \sqrt{\varepsilon_0} = \delta_1 \sqrt{\varepsilon_1} \)

and thus allows

\[ \mathcal{O}(M) = \mathcal{O}_{\varepsilon_0, \delta_0}(M) \geq \mathcal{O}_{\varepsilon_1, \delta_1}(M). \] (E.76)

In the remainder of this section, we carry out the argument described above. Lemma E.29 constructs the aforementioned certificate \( g_1 \) for the constant function \( \zeta_1. \) Lemma E.31 combines this construction with a certificate \( g \) for \( \zeta \) to give a (loose) approximate certificate, for the kernel \( \Theta. \) Theorem E.32 amplifies this construction to reduce the approximation error to an appropriate level. Finally, we finish by incorporating the density \( \rho(x) \) to prove our main result on certificates, Theorem D.2.
Lemma E.29. Let $\zeta_1 \equiv 1$ denote the constant function over $\mathcal{M}$. When $L > C$ and the conditions of Theorem E.1 are satisfied for $\varepsilon = \varepsilon_1 = \frac{31}{100}$, then $g_1 = g_{\varepsilon_1}[\zeta_1]$ satisfies
\[
g_1 \in \Phi \left( \frac{C'}{n \log L} \|\zeta_1\|_{L^2, \varepsilon_1} \right)
\]
and
\[
\Theta^\circ [g_1] - \zeta_1 \in \Phi \left( C'' \|\zeta_1\|_{L^2, \varepsilon_1} \right).
\]
We also have
\[
\mathbb{I}[g_1] \geq \left( 2 - \frac{1}{\varepsilon_1} \right) \frac{\|\zeta_1\|_{L^1}}{C'' n \log (L)}, \tag{E.77}
\]
where $C, C', C''$ and $C'''$ are positive numerical constants.

Proof. Applying Theorem E.1, as conditions of Theorem E.1 for $\varepsilon = \varepsilon_1$ is satisfied, we know $P_{S_{\varepsilon_1}} \tilde{M}_{\varepsilon_1} P_{S_{\varepsilon_1}}$ is invertible over $S_{\varepsilon_1}$. Noting that $\zeta_1$ is a constant function and thus $\zeta_1 \in S_{\varepsilon_1}$, we set
\[
g_1 = g_{\varepsilon_1}[\zeta_1]
\]
\[
= \sum_{\ell=0}^{\infty} (-1)^{\ell} \left( \left( P_{S_{\varepsilon_1}} \tilde{M}_{\varepsilon_1} P_{S_{\varepsilon_1}} \right)^{-1} P_{S_{\varepsilon_1}} \left( \Theta^\circ - \tilde{M}_{\varepsilon_1} \right) P_{S_{\varepsilon_1}} \right) \left( P_{S_{\varepsilon_1}} \tilde{M}_{\varepsilon_1} P_{S_{\varepsilon_1}} \right)^{-1} \zeta_1.
\]
Since $\zeta_1$ is a constant function, all its derivatives are zero and thus $\zeta_1 \in \Phi (\|\zeta_1\|_{L^2}, \varepsilon_1)$. Applying Lemma E.25, we have that
\[
g_1 \in \Phi \left( \frac{C_{\varepsilon_1}}{n \log L} \|\zeta_1\|_{L^2, \varepsilon_1} \right)
\]
for certain $C_{\varepsilon_1} > 0$. The condition $\Theta^\circ [g_1] - \zeta_1 \in \Phi (C'' \|\zeta_1\|_{L^2, \varepsilon_1})$ follows from Lemma E.27 directly.

To control $\mathbb{I}[g_1]$, notice that because $P_{S_{\varepsilon_1}} \tilde{M}_{\varepsilon_1} P_{S_{\varepsilon_1}}$ is an invariant operator stably invertible in $S_{\varepsilon_1}$, and $\zeta_1 \in S_{\varepsilon_1}$, we can set
\[
\tilde{g}_1 = \left( P_{S_{\varepsilon_1}} \tilde{M}_{\varepsilon_1} P_{S_{\varepsilon_1}} \right)^{-1} \zeta_1 = \left( \int_{s = -r_{\varepsilon_1}}^{r_{\varepsilon_1}} \psi^\circ (|s|) ds \right)^{-1} \zeta_1
\]
which is also a positive constant function. We have
\[
g_1 = \sum_{\ell=0}^{\infty} (-1)^{\ell} \left( \left( P_{S_{\varepsilon_1}} \tilde{M}_{\varepsilon_1} P_{S_{\varepsilon_1}} \right)^{-1} P_{S_{\varepsilon_1}} \left( \Theta^\circ - \tilde{M}_{\varepsilon_1} \right) P_{S_{\varepsilon_1}} \right) \tilde{g}_1
\]
\[
= \tilde{g}_1 + \sum_{\ell=1}^{\infty} (-1)^{\ell} \left( \left( P_{S_{\varepsilon_1}} \tilde{M}_{\varepsilon_1} P_{S_{\varepsilon_1}} \right)^{-1} P_{S_{\varepsilon_1}} \left( \Theta^\circ - \tilde{M}_{\varepsilon_1} \right) P_{S_{\varepsilon_1}} \right) \tilde{g}_1,
\]
and from Lemma E.2, we have
\[
\left\| \left( P_{S_{\varepsilon_1}} \tilde{M}_{\varepsilon_1} P_{S_{\varepsilon_1}} \right)^{-1} P_{S_{\varepsilon_1}} \left( \Theta^\circ - \tilde{M}_{\varepsilon_1} \right) P_{S_{\varepsilon_1}} \right\|_{L^2 \rightarrow L^2} \leq 1 - \varepsilon_1,
\]
and so
\[
\|g_1 - \tilde{g}_1\|_{L^1} \leq \sqrt{\text{len}(\mathcal{M})} \|g_1 - \tilde{g}_1\|_{L^2}
\]
\[
\leq \sqrt{\text{len}(\mathcal{M})} \sum_{\ell=1}^{\infty} (1 - \varepsilon_1)^{\ell} \|\tilde{g}_1\|_{L^2}
\]
\[
= \sqrt{\text{len}(\mathcal{M})} \left( \frac{1}{\varepsilon_1} - 1 \right) \|\tilde{g}_1\|_{L^2}
\]
\[
= \left( \frac{1}{\varepsilon_1} - 1 \right) \|\tilde{g}_1\|_{L^1}
\]
The first inequality comes from the equivalence of norms, and the last two lines come from the fact that \( \hat{g}_1 \) is a positive constant function. Thus, following Lemma F.9, there exist constants \( C, C' \) such that

\[
\mathbb{I}[g_1] \geq \mathbb{I}[\hat{g}_1] - \|g_1 - \hat{g}_1\|_{L^1}.
\]

\[
= \left(2 - \frac{1}{\varepsilon_1}\right) \mathbb{I}[\hat{g}_1] - \left(2 - \frac{1}{\varepsilon_1}\right) \left(\int_{r_{\varepsilon_1}}^{r_{\varepsilon_1}} \psi(|s|)ds\right)^{-1} ||\zeta_1||_{L^1}
\]

\[
\geq \left(2 - \frac{1}{\varepsilon_1}\right) \left(||\zeta_1||_{L^1}\right)
\]

\[
\geq \frac{\frac{3\pi n}{8} C'' \log \left(\frac{1 + \frac{1}{3\pi} (L - 3) r_{\varepsilon_1}}{1 + \frac{1}{3\pi} (L - 3) r_{\varepsilon_1}}\right)}{C n \log (L)}
\]

as claimed. \( \blacksquare \)

**Lemma E.30.** Suppose that the conditions of Theorem E.1 are satisfied for both \( \varepsilon = \varepsilon_0 = \frac{1}{20} \) and \( \varepsilon = \varepsilon_1 = \frac{1}{100} \), and let \( g_1, \zeta_1 \) be as in Lemma E.29. Let \( \zeta \in S_{\varepsilon_0} \), and \( g = g_{\varepsilon_0}[\zeta] \). Then

\[
\alpha = -\frac{\psi(\pi) \mathbb{I}[g]}{\psi(\pi) \mathbb{I}[g_1] + 1},
\]

satisfies

\[
|\alpha| \leq \frac{C ||\zeta||_{L^2}}{||\zeta_1||_{L^2}},
\]

where \( C \) is a numerical constant.

**Proof.** Set

\[
\hat{g} = \left(P_{S_{\varepsilon_0}} \bar{M}_{\varepsilon_0} P_{S_{\varepsilon_0}}\right)^{-1} \zeta
\]

again, we have

\[
g = \sum_{\ell=0}^{\infty} (-1)^{\ell} \left(\left(P_{S_{\varepsilon_0}} \bar{M}_{\varepsilon_0} P_{S_{\varepsilon_0}}\right)^{-1} P_{S_{\varepsilon_0}} \left(\Theta^\varepsilon - \bar{M}_{\varepsilon_0}\right) P_{S_{\varepsilon_0}}\right) \hat{g}.
\]

Then there exist constants \( c, C > 0 \) that

\[
\|g\|_{L^1} \leq \sqrt{\text{len}(\mathcal{M})} \|g\|_{L^2}
\]

\[
\leq \sqrt{\text{len}(\mathcal{M})} \sum_{\ell=0}^{\infty} (1 - \varepsilon_0)^{\ell} \|\hat{g}\|_{L^2}
\]

\[
= \sqrt{\text{len}(\mathcal{M})} \frac{1}{\varepsilon_0} \|\hat{g}\|_{L^2}
\]

\[
\leq \sqrt{\text{len}(\mathcal{M})} \frac{1}{\varepsilon_0 \lambda_{\text{min}} (P_{S_{\varepsilon_0}} \bar{M}_{\varepsilon_0} P_{S_{\varepsilon_0}})} \|\zeta\|_{L^2}
\]

\[
\leq \sqrt{\text{len}(\mathcal{M})} \frac{C ||\zeta||_{L^2}}{n \log L},
\]
where in the penultimate inequality, we have used Lemma E.6. Applying the previous lemma, we obtain
\[ |\alpha| \leq \frac{\psi(\pi) [g]}{\psi(\pi) [g_1]} \leq \frac{C n \log(L) \|g\|_L^1}{(2 - \frac{1}{\epsilon_1}) \|\zeta\|_L^1} \leq \frac{C \sqrt{\text{len}(\mathcal{M})} \|\zeta\|_L^2}{\|\zeta_1\|_L^1} = \frac{C \|\zeta\|_L^2}{\|\zeta_1\|_L^2}, \]
where in the final equation we have used that the constant function \( \zeta_1 \) satisfies \( \|\zeta_1\|_L^1 = \sqrt{\text{len}(\mathcal{M})} \|\zeta_1\|_L^2 \).

**Lemma E.31.** Suppose that the conditions of Theorem E.1 are satisfied for both \( \varepsilon = \varepsilon_0 = \frac{1}{20} \) and \( \varepsilon = \varepsilon_1 = \frac{51}{100} \), and \( L \geq (e^{1/2} \min\{\text{len}(\mathcal{M}_+), \text{len}(\mathcal{M}_-)\}/(12\pi^2))^{-\frac{2\pi + 1}{4\pi}} \) for both \( \varepsilon = \varepsilon_0 \) and \( \varepsilon = \varepsilon_1 \). There exist numerical constants \( C, C' > 0 \), such that for every \( \zeta \in \mathcal{S}_{\varepsilon_0} \cap \Phi(C\zeta, \varepsilon_1) \), there exists \( h \) such that
\[ \|h\|_L^1 \leq \frac{C \|\zeta\|_L^2}{n \log L}, \quad (E.78) \]
\[ \|\Theta[h] - \zeta\|_L^2 \leq \frac{P(M_3, M_4, M_5, \text{len}(\mathcal{M}), \Delta_{\varepsilon_0}^{-1}) \times r^3_{\varepsilon_1}}{\log L}, \quad (E.79) \]
\[ \|P_{S_{\varepsilon_0}}[\Theta[h] - \zeta]\|_L^2 \leq \frac{CC \cdot P(M_3, M_4, M_5, \text{len}(\mathcal{M}), \Delta_{\varepsilon_0}^{-1}) \times L^{-\frac{3}{2}}}{\log L}, \quad (E.80) \]
where \( P \) is a polynomial poly\{\( M_3, M_4, M_5, \text{len}(\mathcal{M}), \Delta_{\varepsilon_0}^{-1} \)\} of degree \( \leq 9 \), with degree \( \leq 3 \) in \( M_3, M_4, M_5 \text{len}(\mathcal{M}) \), and degree \( \leq 6 \) in \( \Delta_{\varepsilon_0}^{-1} \). Furthermore, we have
\[ P_{S_{\varepsilon_0}}[\Theta[h] - \zeta] \in \Phi(C' \|\zeta\|_L^2, \varepsilon_0). \quad (E.81) \]

**Proof.** Recall that \( \Theta = \Theta^0 + \psi(\pi) \mathbb{1} \), and let \( g_1 \) denote the solution for \( \zeta_1 = 1 \) as in Lemma E.29. Set \( h = g + \alpha g_1 \), where \( g = g_{\varepsilon_0}[\zeta] \), and
\[ \alpha = -\frac{\psi(\pi) \mathbb{1}[g]}{\psi(\pi) \mathbb{1}[g_1] + 1}. \]
Using Theorem E.1 to control the norms of \( g \) and \( g_1 \), and using Lemma E.30 to control \( |\alpha| \), we have
\[ \|h\|_L^2 \leq \|g\|_L^2 + |\alpha| \|g_1\|_L^2 \leq \frac{C \|\zeta\|_L^2}{n \log L} + |\alpha| \frac{C \|\zeta_1\|_L^2}{n \log L} \leq \frac{C \|\zeta\|_L^2}{n \log L}, \]
establishing (E.78).

From our choice of \( \alpha \),
\[ \Theta[h] - \zeta = \Theta^0[g] - \zeta + \alpha \Theta^0[g_1] + \psi(\pi) \mathbb{1}[g] + \alpha \psi(\pi) \mathbb{1}[g_1] = \Theta^0[g] - \zeta + \alpha (\Theta^0[g_1] - \zeta_1). \quad (E.82) \]
Using Lemma E.28 and the fact that \( \zeta_1 \in \Phi(\|\zeta_1\|, \varepsilon_1) \) we have
\[ \|\Theta[h] - \zeta\|_L^2 \leq \|\Theta^0[g] - \zeta\|_L^2 + |\alpha| \|\Theta^0[g_1] - \zeta_1\|_L^2 \leq C_{\varepsilon_0} C \frac{\|\zeta_1\|_L^2}{\log L} \left( P_3 + \frac{P_1 P_2}{\log L} + \frac{P_3^3}{\log^2 L} \right), \]

54
where $P_{S_{e_0}}$ is a polynomial of $M_3, M_4, M_5, \text{len}(M), \Delta_{\epsilon_0}^{-1}$ of degree less than 9, with degree less than 3 in $M_3, M_4, M_5 \text{len}(M)$, and degree less than 6 in $\Delta_{\epsilon_0}^{-1}$. Here $P_1, P_2$ and $P_3$ are polynomials defined in Lemma E.18, Lemma E.19 and Lemma E.20. This together with (E.83) gives us (E.79).

To obtain the tighter bound (E.80) on $P_{S_{e_0}}(\Theta[h] - \zeta)$, we begin by applying Lemma E.27 with $\zeta \in \Phi(C_{\zeta_1}, \epsilon_0)$ and $\zeta_1 \in \Phi(\|\zeta_1\|_{L^2}, \epsilon_1)$, we have

$$\Theta^0[g] - \zeta \in \Phi\left(C'_{e_0} C_{\zeta_1}, \epsilon_0\right),$$

$$\Theta^0[g] - \zeta_1 \in \Phi\left(C'_{e_1} \|\zeta_1\|_{L^2}, \epsilon_1\right) \subseteq \Phi\left(C'_{e_1} \|\zeta_1\|_{L^2}, \epsilon_0\right)$$

for certain $C'_{e_0}, C'_{e_1} > 0$. Using $\Theta[h] - \zeta = \Theta^0[g] - \zeta + \alpha(\Theta^0[g_1] - \zeta_1)$, we have

$$C'_{e_0} C_{\zeta} + C'_{e_1} \|\zeta_1\|_{L^2} \times C'_{e_1} \|\zeta_1\|_{L^2} \leq C'' C_{\zeta},$$

for some constant $C'' > 0$ and thus

$$\Theta[h] - \zeta \in \Phi(C'' C_{\zeta}, \epsilon_0).$$

Applying Lemma E.28 with $w = \Theta[h] - \zeta$ and simplifying with (E.84) we obtain

$$\|P_{S_{e_0}}(\Theta[h] - \zeta)\|_{L^2} \leq C''' C_{\zeta} \frac{P\left(M_3, M_4, M_5, \text{len}(M), \Delta_{\epsilon_0}^{-1}\right) \times r_{e_0}^3}{\log L}.$$  

(E.80) follows from $r_{e_0} < 6\pi L^{-4/9}$, which implies that $r_{e_0}^3/\log L \leq L^{-4/3}$ when $L$ is larger than an appropriate numerical constant.

Finally, since $P_{S_{e_0}} \Theta^0[g] = P_{S_{e_0}} \zeta$,

$$P_{S_{e_0}}(\Theta[h] - \zeta) = \alpha P_{S_{e_0}}(\Theta^0[g] - \zeta_1).$$  

(E.87)
From Lemma E.22, \( P_{S_{\epsilon_0}} \) commutes with differentiation, and so for any \( i \)-times differentiable \( w \),
\[
\| (P_{S_{\epsilon_0}} w^{(i)}) \|_{L^2} \leq \| w^{(i)} \|_{L^2}.
\]
Applying this with \( i = 0, 1, 2, 3 \), we see that for any \( w \in \Phi(C_w, \epsilon) \), \( P_{S_{\epsilon_0}} w \in \Phi(C_w, \epsilon) \). Applying this observation to (E.85), we have
\[
P_{S_{\epsilon_0}} (\Theta g_1 - \zeta_1) \in \Phi(C'_{\epsilon_1} \| \zeta_1 \|_{L^2}, \epsilon_0).
\]
Combining with (E.87), we have
\[
P_{S_{\epsilon_0}} (\Theta [h] - \zeta) \in \Phi\left( \| \alpha C'_1 \|_{L^2}, \epsilon_0 \right) \subseteq \Phi\left( C' C_{\epsilon_1} \| \zeta \|_{L^2}, \epsilon_0 \right),
\]
which is (E.81). Here, we have used the bound on \( \alpha \) from the previous lemma. This completes the proof. 

**Theorem E.32 (Certificates for DC kernel).** Suppose that the conditions of Theorem E.1 are satisfied for both \( \epsilon = \epsilon_0 = \frac{1}{10} \) and \( \epsilon = \epsilon_1 = \frac{51}{100} \) and \( L \geq \left( \epsilon_1^{1/2} \min\{\text{len}(M_+), \text{len}(M_-)\}/(12\pi^2) \right)^{-\frac{n+1}{2}} \) for both \( \epsilon = \epsilon_0 \) and \( \epsilon = \epsilon_1 \). There exist constants \( C, C'' \) such that for any number \( K > 0 \), when
\[
L \geq CK^4P\left(M_3, M_4, M_5, \text{len}(M), \Delta^{-1}_0\right)^4,
\]
then for any \( \zeta \in S_{\epsilon_0} \cap \Phi(K \| \zeta \|_{L^2}, \epsilon_0) \) there exists a certificate \( h \) satisfying
\[
\| \Theta [h] - \zeta \|_{L^2} \leq \| \zeta \|_{L^2} L^{-1},
\]
with
\[
\| h \|_{L^2} \leq \frac{C'' \| \zeta \|_{L^2}}{n \log L}.
\]
In (E.88), \( P \) is the polynomial defined in Lemma E.31.

**Proof.** Let \( \zeta(0) = \zeta \), and iteratively define
\[
\zeta(i+1) = -P_{S_{\epsilon_0}} (\Theta [h(i)] - \zeta(i)) \in S_{\epsilon_0},
\]
where \( h(i) \) is the approximate certificate of \( \zeta(i) \in S_{\epsilon_0} \) constructed in Lemma E.31. From (E.81), we have
\[
\zeta(i+1) \in \Phi\left( C_1 \| \zeta(i) \|_{L^2}, \epsilon_0 \right),
\]
where \( C_1 \) is a numerical constant. Hence, for \( i \geq 1 \), from Lemma E.31 we have
\[
\| h(i) \|_{L^2} \leq \left( \frac{C_2}{n \log L} \right) \| \zeta(i) \|_{L^2},
\]
\[
\| \Theta [h(i)] - \zeta(i) \|_{L^2} \leq C_2 C_1 \| \zeta(i-1) \|_{L^2} \frac{P \epsilon_{\epsilon_1}}{\log L},
\]
\[
\| P_{S_{\epsilon_0}} (\Theta [h(i)] - \zeta(i)) \|_{L^2} \leq C_2 C_1 \| \zeta(i-1) \|_{L^2} P L^{-4/3}.
\]
For \( i = 0 \), as \( \zeta(0) \in \Phi\left( K \| \zeta \|_{L^2}, \epsilon_0 \right) \), this simplifies to
\[
\| h(0) \|_{L^2} \leq \left( \frac{C_2}{n \log L} \right) \| \zeta \|_{L^2},
\]
\[
\| \Theta [h(0)] - \zeta(0) \|_{L^2} \leq C_2 K \| \zeta \|_{L^2} \frac{P \epsilon_{\epsilon_1}}{\log L},
\]
\[
\| P_{S_{\epsilon_0}} (\Theta [h(0)] - \zeta(0)) \|_{L^2} \leq C_2 K \| \zeta \|_{L^2} P L^{-4/3}.
\]
We use these relationships to control \( \| \zeta(i) \|_{L^2} \). As \( r_{\epsilon_1} \leq 6\pi L^{-1/10} \), there exists a constant \( C \) such that when \( L \geq CK^4P^4, C_2 K \frac{P \epsilon_{\epsilon_1}}{\log L} \leq \tau = \frac{1}{2} \) and \( C_2 C_1 \frac{P \epsilon_{\epsilon_1}}{\log L} \leq \tau^2 \). We argue by induction that
\[
\| \zeta(i) \|_{L^2} \leq \tau^i \| \zeta \|_{L^2} \quad \forall \ i \geq 0.
\]
We next verify that \( \Theta \). We set \( k \), where \( i \) such that \( \tau^i \| \zeta \| L^2 \). This is true by construction for \( i = 0 \), while for \( i = 1 \) it follows from (E.91). Finally, for \( i \geq 2 \), using (E.90) and inductive hypothesis, we have
\[
\| \zeta \| L^2 \leq \| \Theta [h_{i-1}] - \zeta(i-1) \| L^2 \leq C_2 C_1 \| \zeta(i-2) \| L^2 \leq \tau^2 \| \zeta(i-2) \| L^2 \leq \tau^i \| \zeta \| L^2 ,
\]
as claimed.
We set
\[
h = \sum_{i=0}^{k} h_{i(i)},
\]
where \( k \) will be specified below. By construction,
\[
\| h \| L^2 \leq \sum_{i=0}^{k} \| h_{i(i)} \| L^2 \leq \frac{C_1}{n \log L} \sum_{i=0}^{k} \| \zeta(i) \| L^2 \leq \frac{2C_1}{n \log L} \| \zeta \| L^2 ,
\]
as claimed.

We next verify that \( \Theta [h] \) is an accurate approximation to \( \zeta \):
\[
\| \Theta [h] - \zeta(0) \| L^2 \leq \| P_{S_{e0}}[\Theta [h] - \zeta(0)] \| L^2 + \| P_{S_{e0}} \Theta [h] \| L^2 \leq \| P_{S_{e0}} [\Theta [h] - \zeta(0)] \| L^2 + \sum_{i=0}^{k} \| P_{S_{e0}} \Theta [h_{i(i)}] \| L^2 \leq \tau^{k+1} \| \zeta(0) \| L^2 + C_2 C_1 P L^{-4/3} \sum_{i=1}^{k} \| \zeta(i-1) \| L^2 \leq \tau^{k+1} \| \zeta(0) \| L^2 + C_2 K P L^{-4/3} \| \zeta \| L^2 \leq \tau^{k+1} \| \zeta(0) \| L^2 + C_2 P L^{-4/3} (2C_1 + K) \| \zeta \| L^2 .
\]
Choosing \( k \) appropriately, and ensuring that \( C_2(2C_1 + K) P L^{-4/3} < \frac{1}{2} L^{-1} \), establishes (E.89). The latter condition follows immediately, from \( L \geq CK^3 P^4 \) for appropriately large \( C \).

**Lemma E.33.** There exist constants \( C, C', C'', C''' \) and a polynomial \( P = \text{poly} \{ M_3, M_4, M_5, \text{len}(M) , \Delta^{-1} \} \) of degree \( \leq 36 \), with degree \( \leq 12 \) in \( M_3, M_4, M_5 \), \( \text{len}(M) \), and degree \( \leq 24 \) in \( \Delta^{-1} \) such that for any number \( K > 0 \), when
\[
L \geq \max \left\{ \exp(C' \text{len}(M) K), \left( \frac{1}{\Delta \sqrt{1+K^2}} \right)^{C''' \Theta(M)}, C'' K^{10}, K^4 P, P_{\text{max}}^{12} \right\}
\]
then for any real \( \zeta \in \Phi(K \| \zeta \| L^2, \frac{1}{20}) \), there exists a real certificate \( g : \mathcal{M} \to \mathbb{R} \) with
\[
\| g \| L^2 \leq \frac{C \| \zeta \| L^2}{\rho_{\text{min}} n \log L}
\]
such that
\[
\| \Theta_{\mu} [g] - \zeta \| L^2 \leq \| \zeta \| L \leq L^{-1}
\]
We first show that there exists constants $C, C', C''$ that under condition of the lemma, conditions of Theorem E.32 are satisfied. From definition, $\Delta = \Delta_{\varepsilon_0} \leq \Delta_{\varepsilon_1}$, as we choose $\delta_0 = 1 - \varepsilon_0, \delta_1 = \delta_0 \sqrt{\varepsilon_0} / \sqrt{\varepsilon_1} < 1 - \varepsilon_1$, we have $\delta_0 \sqrt{\varepsilon_0} = \delta_1 \sqrt{\varepsilon_1}$ and thus $\mathcal{O}(\varepsilon) = \mathcal{O}_{\varepsilon_0, \varepsilon_1}(\varepsilon) \geq \mathcal{O}_{\varepsilon_1, \varepsilon_1}(\varepsilon)$ as in (E.76). Thus conditions of Theorem E.1 for $\varepsilon = \varepsilon_1$ can be absorbed into conditions for $\varepsilon = \varepsilon_0$ with a change of constant factors.

Furthermore, as $\Delta \leq \frac{\sqrt{\varepsilon_0}}{\kappa} < \frac{1}{2 \sqrt{1 + \kappa}}$, we can alter the condition in Theorem E.1 by

$$
\left(1 + \frac{1}{\Delta \sqrt{1 + \kappa}} \right)^{C''} \mathcal{O}(\varepsilon) \leq \left( \frac{1}{\Delta \sqrt{1 + \kappa}} \right) \text{ and thus reduce to the form in the statement.}
$$

We then notice that from Lemma E.3 $\min\{\text{len}(\mathcal{M}_+), \text{len}(\mathcal{M}_-)\} \geq \frac{1}{2}$, and we can choose $C'$ such that $\exp(C' \text{ len}(\mathcal{M}) \kappa) \geq \exp(C') \geq C_{\varepsilon_0}$. Similarly, by choosing $C''\kappa$ appropriately, $L \geq C''\kappa^{10}$ implies both $L \geq (e^{1/26\pi} \kappa)^{\frac{a_{\varepsilon_1}}{2}}$ and $L \geq (e^{-1/2} \varepsilon_0 \sqrt{\kappa})^{\frac{a_{\varepsilon_0}}{2}}$ for both $\varepsilon = \varepsilon_0$ and $\varepsilon = \varepsilon_1$ as $a_{\varepsilon_0} > \frac{1}{\varepsilon_0}$ and $a_{\varepsilon_1} > \frac{1}{\varepsilon_1}$.

From Theorem E.32, we know there exists $g$ such that

$$
\|\Theta[h] - \zeta\|_{L^2} \leq \|\zeta\|_{L^\infty} L^{-1} \tag{E.92}
$$

with

$$
\|g\|_{L^2} \leq \frac{C\|\zeta\|_{L^2}}{n \log L}. \tag{E.93}
$$

We can further require this $g$ to be a real function over the manifold. To see this, notice that for any $x, x' \in \mathcal{M}$, both the kernel $\Theta(x, x')$ and $\zeta(x)$ are real, thus if we take the real component of $g$ as $\bar{g} = (g + \bar{g})/2$, then we have $\|\bar{g}\|_{L^2} \leq \|g\|_{L^2}$ and further by the triangle inequality

$$
|\Theta[\bar{g}](x) - \zeta(x)| = \left| \int_{x' \in \mathcal{M}} \Theta(x, x') \bar{g}(x') dx' - \zeta(x) \right|
\leq \left( \int_{x' \in \mathcal{M}} |\Theta(x, x') \bar{g}(x') dx'|^2 \right)^{1/2}
\leq \left( \int_{x' \in \mathcal{M}} |\Theta(x, x')(1/2)(g - \bar{g})(x') dx'|^2 \right)^{1/2}
= \int_{x' \in \mathcal{M}} \Theta(x, x') |g(x') dx' - \zeta(x)|.
$$

The last line comes from the fact that $\int_{x' \in \mathcal{M}} \Theta(x, x') \bar{g}(x') dx' - \zeta(x)$ and $\int_{x' \in \mathcal{M}} \Theta(x, x')(1/2)(g - \bar{g})(x') dx'$ are the pure real and imaginary part of $\int_{x' \in \mathcal{M}} \Theta(x, x') g(x') dx' - \zeta(x)$. Thus $\bar{g}$ is real and also satisfies (E.92) and (E.93).

To include the density, define $g_\mu(x) = g(x)/\rho(x)$. We get

$$
\|g_\mu\|_{L^2_{\rho}}^2 = \int_{x \in \mathcal{M}} |g_\mu(x)|^2 \rho(x) dx
= \int_{x \in \mathcal{M}} |g(x)|^2 \rho^{-1}(x) dx
\leq \min_{x' \in \mathcal{M}} \rho^{-1}(x') \int_{x \in \mathcal{M}} |g(x)|^2 dx
= \rho^{-1}_{\min} \|g\|_{L^2}^2.
$$

Then for $\zeta$, we have

$$
\|\zeta\|_{L^2_{\rho}}^2 = \int_{x \in \mathcal{M}} |\zeta(x)|^2 \rho(x) dx
\geq \rho_{\min} \|\zeta\|_{L^2}^2.
$$

This gives

$$
\|g_\mu\|_{L^2_{\rho}} \leq \rho^{-1/2}_{\min} \|g\|_{L^2}.
$$
We will prove that 
\[ \hat{\psi} \leq \rho_{\min}^{-1/2} C \| \zeta \|_{L^2} \]
\[ \leq \frac{C \| \zeta \|_{L^2}}{\rho_{\min} n \log L}. \]

On the other hand, we have
\[ \| \Theta_{\mu}[g] - \zeta \|_{L^2}^2 = \int_{x \in \mathcal{M}} \left( \int_{x' \in \mathcal{M}} \Theta(x, x')g(x)\rho(x')dx' - \zeta(x) \right)^2 \rho(x)dx \]
\[ \leq \max_{x'' \in \mathcal{M}} \rho(x'') \int_{x \in \mathcal{M}} \left( \int_{x' \in \mathcal{M}} \Theta(x, x')g(x')dx' - \zeta(x) \right)^2 dx \]
\[ = \rho_{\max} \| \Theta[g] - \zeta \|_{L^2}. \]

Notice that \( P' = CP + \text{len}(\mathcal{M})^3 \) is still a polynomial of degree 3 in \( \text{len}(\mathcal{M}) \). Then when \( L \geq \max\{ P'^4, \rho_{\max}^2 \} \), we get
\[ \| \Theta_{\mu}[g] - \zeta \|_{L^2}^2 \leq \rho_{\max}^{1/2} C P L^{-4} \| \zeta \|_{L^2} \]
\[ = L^{-1} \| \zeta \|_{L^\infty} \left( \frac{CP}{L^{1/4}} \frac{\rho_{\max}^{1/2}}{L^{1/24}} \right) \]
\[ \leq L^{-1} \| \zeta \|_{L^\infty}. \]

As \( P \) in Lemma E.31 is a polynomial \( \{M_3, M_4, M_5, \text{len}(\mathcal{M}), \Delta^{-1}\} \) of degree \( \leq 9 \), with degree \( \leq 3 \) in \( M_2, M_4, M_5 \) len\( (\mathcal{M}) \), and degree \( \leq 6 \) in \( \Delta^{-1} \), we have \( P'^4 \) is of the right degree requirement as in the statement of the theorem.

**F  Bounds for the Skeleton Function \( \psi \)**

In this section, we are going to provide sharp bounds on the “skeleton” function \( \psi \) and its higher-order derivatives. We recall that the angle evolution function is defined as
\[ \varphi(t) = \arccos \left( \frac{t}{\pi} \cos t + \frac{1}{\pi} \sin t \right), \quad t \in [0, \pi]. \]

Define \( \varphi^{[0]} = \text{Id}, \varphi^{[\ell]} \) as \( \varphi \)’s \( \ell \)-fold composition with itself (which will be referred to as the iterated angle evolution function). Then the skeleton is defined as
\[ \psi(t) = \frac{n}{2} \sum_{\ell=0}^{L-1} \xi_{\ell}(t), \]
where
\[ \xi_{\ell}(t) = \prod_{\ell' = \ell}^{L-1} \left( 1 - \frac{1}{\pi} \varphi^{[\ell']}(t) \right), \quad \ell = 0, \ldots, L - 1. \]

To analyze the function \( \psi \), we will establish in this section several “sharp-modulo-constants” estimates that connect \( \psi \) to a much simpler function, derived using the local behavior of \( \varphi \) at 0 and its consequences for the iterated compositions \( \varphi^{[\ell]} \) that appear in the definition of \( \psi \). In particular, let us define \( \tilde{\varphi} : [0, \pi] \to [0, \pi] \) by \( \tilde{\varphi}(t) = t/(1 + t/(3\pi)) \), so that
\[ \tilde{\varphi}^{[\ell]}(t) = \frac{t}{1 + \ell t/(3\pi)}, \]
and moreover define
\[ \tilde{\xi}_{\ell}(t) = \prod_{\ell' = \ell}^{L-1} \left( 1 - \frac{\tilde{\varphi}^{[\ell']}(t)}{\pi} \right), \quad \tilde{\psi}(t) = \frac{n}{2} \sum_{\ell=0}^{L-1} \tilde{\xi}_{\ell}(t). \]

We will prove that \( \tilde{\varphi}^{[\ell]} \) provides a sharp approximation to \( \varphi^{[\ell]} \) (Lemmas F.2 and F.3), and then work out a corresponding sharp approximation of \( \tilde{\psi} \) to \( \psi \) (Lemmas F.7 and F.9). We will then derive
estimates for the low-order derivatives of $\psi$ in Appendix F.4. Unfortunately, it is impossible to obtain $L^1$ estimates for $\psi$ in terms of $\hat{\psi}$ that are sharp enough to facilitate operator norm bounds for $\Theta_\mu$, which would let us construct certificates for an operator with kernel $\hat{\psi}$ rather than the NTK $\Theta_\mu$; but the estimates we derive in this section will be nonetheless sufficient to enable our localization and certificate construction arguments in Appendix E.

We note that bounds similar to a subset of the bounds in this section have been developed in an $L$-asymptotic, large-angle setting by [5]. The bounds we develop here are non-asymptotic and hold for all angles, and are established using elementary arguments that we believe are slightly more transparent. We reuse (and restate in Appendix G) some estimates from [6, Section C] here, but the majority of our estimates will be fundamentally improved (a representative example is Lemma F.3).

Throughout this section, we use $\dot{\phi}$, $\ddot{\phi}$, and $\dddot{\phi}$ to represent first, second, and third derivatives of $\phi$ (see Lemma G.5 for basic regularity assertions for this function and its iterated compositions) and likewise for $\xi$ and $\psi$. In particular, for example, in our notation the function $\dot{\phi}[\ell]$ refers to the derivative of $\phi[\ell]$, not the $\ell$-fold iterated composition of $\dot{\phi}$. Although this leads to an abuse of notation, the concision it enables in our proofs will be of use.

### F.1 Sharp Lower Bound for the Iterated Angle Evolution Function

**Lemma F.1.** One has

$$\varphi(t) \leq \frac{t}{1 + t/(3\pi)}, \quad t \in [0, \pi].$$

**Proof.** As $\cos$ is monotonically decreasing in $[0, \pi)$, it is the same as proving

$$(1 - \frac{t}{\pi}) \cos(t) + \frac{\sin t}{\pi} - \cos \frac{t}{1 + \frac{t}{3\pi}} \geq 0$$

We have the gradient as

$$-(1 - \frac{t}{\pi}) \sin t + \sin \left(\frac{t}{1 + t/(3\pi)}\right) \frac{1}{(1 + t/(3\pi))^2}$$

$$\geq -(1 - \frac{t}{\pi}) \sin t + \sin t \frac{1}{(1 + t/(3\pi))^2}$$

$$\geq (1 - \frac{t}{\pi}) + \frac{1}{(1 + \frac{t}{3\pi})^2} \sin t$$

$$\geq 0$$

For the first inequality, we use the estimate

$$\sin(ax) \geq x \sin a; \quad 0 \leq x \leq 1, \quad 0 \leq a \leq \pi,$$

which is easily established using concavity of $\sin$ on $[0, \pi]$ and the secant line characterization, and for the final inequality, we use the estimate $1 - 3a \leq \frac{1}{(1+a)^2}$ for any $a > -1$, which follows from convexity of $a \mapsto (1 + a)^{-3}$ on this domain and the tangent line characterization (at $a = 0$). Since at $t = 0$, we have the inequality holds, we know it holds for the whole interval $[0, \pi]$ by the mean value theorem.

**Lemma F.2** (Corollary of [6, Lemma C.12]). If $\ell \in \mathbb{N}_0$, one has the "fluid" estimate for the iterated angle evolution function

$$\varphi^{[\ell]}(t) \leq \frac{t}{1 + \ell t/(3\pi)}.$$  

**Proof.** Follow the argument of [6, Lemma C.12], but use Lemma F.1 as the basis for the argument instead of Lemma G.4.

**Lemma F.3.** There exists an absolute constant $C_0 > 0$ such that for all $\ell \in \mathbb{N}$

$$\hat{\varphi}^{[\ell]} - \varphi^{[\ell]} \leq C_0 \frac{\log(1 + \ell)}{\ell^2}.$$  

(F.2)
As a consequence, there exist absolute constants $C, C', C'' > 0$ such that for any $0 < \varepsilon \leq 1/2$, if $L \geq C\varepsilon^{-2}$ then for every $t \in [0, C'\varepsilon^2]$ one has

$$\varphi^{[\ell]}(t) \leq \frac{1 + \varepsilon}{(1 + Lt/(3\pi))^2};$$

and for every $t \in [0, \pi]$ one has

$$\bar{\varphi}^{[\ell]}(t) \leq \frac{C''}{(1 + Lt/(3\pi))^2}.$$

Finally, we have for $\ell > 0$

$$\xi_\ell(t) \leq (1 + e^{6C_0} \log(1 + \ell)) \bar{\xi}_\ell(t), \quad (F.3)$$

and if $L \geq 3$

$$\psi(t) \leq \bar{\psi}(t) + 4e^{6C_0} \log^2 L.$$

**Proof.** Fix $L \in \mathbb{N}$ arbitrary. We prove (F.2) first, then use it to derive the remaining estimates. The main tool is an inductive decomposition: start by writing

$$\varphi^{[\ell]}(t) - \varphi^{[\ell]}(t) = \bar{\varphi} \circ \varphi^{[\ell-1]}(t) - \varphi \circ \varphi^{[\ell-1]}(t)$$

$$= \bar{\varphi} \circ \varphi^{[\ell-1]}(t) - \bar{\varphi} \circ \varphi^{[\ell-1]}(t) + \bar{\varphi} \circ \varphi^{[\ell-1]}(t) - \varphi \circ \varphi^{[\ell-1]}(t),$$

and then use the definition of $\bar{\varphi}$ to simplify the first term on the RHS of the final equation (via direct algebraic manipulation) to

$$\bar{\varphi} \circ \varphi^{[\ell-1]}(t) - \bar{\varphi} \circ \varphi^{[\ell-1]}(t) = \frac{\bar{\varphi}^{[\ell-1]}(t) - \varphi^{[\ell-1]}(t)}{(1 + \frac{1}{3\pi} \varphi^{[\ell-1]}(t)) \left(1 + \frac{1}{3\pi} \bar{\varphi}^{[\ell-1]}(t)\right)}.$$

This gives an expression for the difference $\bar{\varphi}^{[\ell]}(t) - \varphi^{[\ell]}(t)$ as an affine function of the previous difference $\bar{\varphi}^{[\ell-1]}(t) - \varphi^{[\ell-1]}(t)$:

$$\bar{\varphi}^{[\ell]}(t) - \varphi^{[\ell]}(t) = \frac{\bar{\varphi}^{[\ell-1]}(t) - \varphi^{[\ell-1]}(t)}{(1 + \frac{1}{3\pi} \varphi^{[\ell-1]}(t)) \left(1 + \frac{1}{3\pi} \bar{\varphi}^{[\ell-1]}(t)\right)} + \bar{\varphi} \circ \varphi^{[\ell-1]}(t) - \varphi \circ \varphi^{[\ell-1]}(t),$$

and unraveling inductively, we obtain

$$\bar{\varphi}^{[\ell]}(t) - \varphi^{[\ell]}(t) = \sum_{\ell=0}^{L-1} \left( \prod_{\ell'=\ell+1}^{L-1} \frac{1}{(1 + \frac{1}{3\pi} \varphi^{[\ell']}(t)) \left(1 + \frac{1}{3\pi} \bar{\varphi}^{[\ell']}(t)\right)} \right) \left( \bar{\varphi} - \varphi \right) \circ \varphi^{[\ell]}(t),$$

where for concision we write $(\bar{\varphi} - \varphi)(t) = \bar{\varphi}(t) - \varphi(t)$. Note that all the product coefficients in this expression are nonnegative numbers. Denoting by $C_1$ the constant attached to $\ell^3$ in the result Lemma F.13 and defining $C_1 = \max \left\{ \tilde{C}_1, 1 \right\}$, Lemma F.13 gives

$$\bar{\varphi}^{[\ell]}(t) - \varphi^{[\ell]}(t) \leq C_1 \sum_{\ell=0}^{L-1} \left( \prod_{\ell'=\ell+1}^{L-1} \frac{1}{(1 + \frac{1}{3\pi} \bar{\varphi}^{[\ell']}(t)) \left(1 + \frac{1}{3\pi} \varphi^{[\ell']}(t)\right)} \right) \varphi^{[\ell]}(t)^3. \quad (F.4)$$

To prove (F.2), we will use a two-stage approach:

1. **(First pass)** First, we will control only the first factor in the product term in (F.4) using Lemma F.13, given that $\varphi \geq 0$ allows us to upper bound by the product term without the second factor. The resulting bound on the LHS of (F.4) will be weaker (in terms of its dependence on $L$) than (F.2).

2. **(Second pass)** After completing this control, we will have obtained a lower bound on $\varphi^{[\ell]}$; we can then return to (F.4) and use this lower bound to get control of both factors in the product term, which will allow us to sharpen our previous analysis and establish the claimed bound (F.2).
First pass. We have

\[ \prod_{\ell' = t + 1}^{L - 1} \frac{1}{1 + \frac{Lt'}{3\pi}} \tilde{\varphi}[\ell'](t) = \frac{1 + (t + 1)t}{1 + \frac{Lt}{3\pi}}. \]

Tossing the product term involving \( \varphi[\ell'] \) and applying Lemma F.2 in (F.4), we thus have a bound

\[ \tilde{\varphi}^{[L]}(t) - \varphi^{[L]}(t) \leq \frac{C_1 t^3}{1 + \frac{Lt}{3\pi}} + \frac{C_1 Lt^3/(3\pi)}{1 + \frac{Lt}{3\pi}} + \frac{C_1 Lt^4/(3\pi)}{1 + \frac{Lt}{3\pi}}. \]

For the first term in this expression, we calculate using an estimate from the integral test

\[ \sum_{\ell = 0}^{L - 1} \frac{t^3}{(1 + \frac{Lt}{3\pi})^2} \leq t^3 + \int_0^L \frac{t^3}{(1 + \frac{Lt}{3\pi})^2} \, dt \]

\[ = t^3 + 3\pi t^2 \left( 1 - \frac{1}{1 + \frac{Lt}{3\pi}} \right) \]

\[ = t^3 + \frac{Lt^3}{1 + \frac{Lt}{3\pi}}, \]

and for the second term, we calculate similarly

\[ \sum_{\ell = 0}^{L - 1} \frac{t^3}{(1 + \frac{Lt}{3\pi})^3} \leq t^3 + \int_0^L \frac{t^3}{(1 + \frac{Lt}{3\pi})^3} \, dt \]

\[ = t^3 + \frac{3\pi t^2}{2} \left( 1 - \frac{1}{(1 + \frac{Lt}{3\pi})^2} \right) \]

\[ = t^3 + \frac{Lt^3}{1 + \frac{Lt}{3\pi}} \]

\[ \leq t^3 + \frac{Lt^3}{1 + \frac{Lt}{3\pi}}. \]

Combining these results gives

\[ \tilde{\varphi}^{[L]}(t) - \varphi^{[L]}(t) \leq \frac{C_1 t^3}{1 + \frac{Lt}{3\pi}} + \frac{C_1 Lt^3}{1 + \frac{Lt}{3\pi}} + \frac{C_1 Lt^4/(3\pi)}{1 + \frac{Lt}{3\pi}} \]

\[ \leq \frac{3\pi C_1 t}{L} \left( 3\pi + 2t + \frac{1}{3\pi} t^2 \right). \]  

(F.5)

This bound gives us a nontrivial estimate as far out as \( t = \pi \), but the result is weaker there than what we need. We can proceed with a bootstrapping approach to improve our result for large angles. To begin, we have shown via (F.5)

\[ \varphi^{[L]}(t) \geq \tilde{\varphi}^{[L]}(t) - \frac{16\pi^2 C_1 t}{L}. \]

Let us write \( t_0 = C/\sqrt{L} \), where \( C > 0 \) is a constant we will optimize below, and define

\[ \tilde{\varphi}_L(t) = \begin{cases} 
\tilde{\varphi}^{[L]}(t) - \frac{16C_1 \pi^2 t}{L} & 0 \leq t \leq t_0 \\
\tilde{\varphi}^{[L]}(t_0) - \frac{16C_1 \pi^2 t_0}{L} & t_0 \leq t \leq \pi.
\end{cases} \]

The notation here is justified by noticing that \( \varphi^{[L]} \) is concave and nondecreasing, so that our previous estimates imply \( \varphi^{[L]} \geq \tilde{\varphi}_L \). It follows

\[ \tilde{\varphi}^{[L]} - \varphi^{[L]} \leq \tilde{\varphi}^{[L]} - \tilde{\varphi}_L. \]

Our previous bound (F.5) is an increasing function of \( t \), and sufficient for \( 0 \leq t \leq t_0 \). For \( t \geq t_0 \), we have

\[ \tilde{\varphi}^{[L]}(t) - \tilde{\varphi}_L(t) \leq \frac{16C_1 \pi^2 t_0}{L} + \tilde{\varphi}^{[L]}(t) - \tilde{\varphi}^{[L]}(t_0), \]

62
and we can calculate using increasingness of $\hat{\varphi}^{[L]}$

$$\hat{\varphi}^{[L]}(t) - \hat{\varphi}^{[L]}(t_0) \leq \frac{\pi}{1 + L/3} - \frac{C}{\sqrt{L} + CL/(3\pi)}$$

$$= \frac{\pi \sqrt{L} - C}{(1 + L/3)(\sqrt{L} + CL/(3\pi))} \leq \frac{9\pi^2}{CL^{3/2}},$$

whence the bound

$$\hat{\varphi}^{[L]} - \varphi^{[L]} \leq \frac{\pi}{L^{5/2}} \left(16\pi C_1 + \frac{9\pi^2}{C}\right) \leq \frac{24\pi^2 \sqrt{C_1}}{L^{3/2}}$$

valid on the entire interval $[0, \pi]$; the final inequality corresponds to the choice $C = \frac{3}{4\sqrt{C_1}}$.

**Second pass.** To start, with an eye toward the unused product term in (F.4), we have from (F.6)

$$1 + \frac{1}{3\pi} \varphi^{[L]}(t) \geq 1 + \frac{1}{3\pi} \hat{\varphi}^{[L]}(t) - \frac{8\pi \sqrt{C_1}}{L^{3/2}}.$$

Using the numerical inequality $e^{-2x} \leq 1 - x$, valid for $0 \leq x \leq 1/2$ at least, we have if $L \geq (256\pi^2 C_1)^{1/3}$

$$1 + \frac{1}{3\pi} \varphi^{[L]}(t) \geq \exp\left(-\frac{16\pi \sqrt{C_1}}{L^{3/2}}\right) \left(1 + \frac{1}{3\pi} \hat{\varphi}^{[L]}(t)\right).$$

Applying this bound to terms in the second product term in (F.4) with index $\ell \geq \left\lceil \left(256\pi^2 C_1\right)^{1/3}\right\rceil \equiv r(C_1)$, we therefore have

$$\hat{\varphi}^{[L]}(t) - \varphi^{[L]}(t) \leq C_1 \sum_{\ell = 0}^{L-1} \left(\varphi^{[\ell]}(t)\right)^3 \left(\prod_{\ell' = \max\left\{r(C_1), \ell+1\right\}}^{L-1} \frac{1}{\left(1 + \frac{1}{3\pi} \hat{\varphi}^{[\ell']}(t)\right)^2}\right)$$

$$\leq C_1 \sum_{\ell = 0}^{L-1} \left(\varphi^{[\ell]}(t)\right)^3 \exp\left(16\pi \sqrt{C_1} \sum_{\ell' = \max\left\{r(C_1), \ell+1\right\}}^{L-1} \frac{1}{(\ell')^{3/2}}\right)$$

$$\times \left(\prod_{\ell' = \max\left\{r(C_1), \ell+1\right\}}^{L-1} \frac{1}{\left(1 + \frac{1}{3\pi} \hat{\varphi}^{[\ell']}(t)\right)^2}\right)$$

$$= C_1 e^{16\pi \sqrt{C_1} \zeta(3/2)} \sum_{\ell = 0}^{L-1} \left(\varphi^{[\ell]}(t)\right)^3 \left(1 + \frac{\max\left\{r(C_1), \ell+1\right\} t}{1 + \frac{4\ell}{3\pi}}\right)^2$$

$$= C_1 e^{16\pi \sqrt{C_1} \zeta(3/2)} \frac{1}{(1 + \frac{L^3}{3\pi})^2} \left(1 + \frac{r(C_1) - 1}{3\pi} t\right)^2 \sum_{\ell=0}^{r(C_1)-2} \left(\varphi^{[\ell]}(t)\right)^3$$

$$+ \sum_{\ell = r(C_1)-1}^{L-1} \left(\varphi^{[\ell]}(t)\right)^3 \left(1 + \frac{(\ell+1) t}{3\pi}\right)^2.$$

Now, since $\varphi \leq \hat{\varphi}$, we have

$$\left(1 + \frac{r(C_1) - 1}{3\pi} t\right)^2 \sum_{\ell=0}^{r(C_1)-2} \left(\varphi^{[\ell]}(t)\right)^3 \leq r(C_1)^2 t^3,$$

\footnote{Although it has a different meaning in our argument at large, here and in some subsequent bounds $\zeta(x) = \sum_{n=1}^{\infty} n^{-x}$ denotes the Riemann zeta function. In this setting, we have $\zeta(3/2) \leq \epsilon$.}
and
\[
\sum_{\ell=r(C_1)}^{L-1} \left( \frac{C_1 e^{16\pi \sqrt{C_1\xi(3/2)}}}{(1 + \frac{Lt}{3\pi})^2} \right)^3 \left( 1 + \frac{(\ell + 1)t}{3\pi} \right)^2 \leq 2t^3 \sum_{\ell=0}^{L-1} \frac{1}{1 + \frac{Lt}{3\pi}} \\
\leq 2t^3 + 2t^3 \int_0^{L} \frac{1}{1 + \frac{Lt}{3\pi}} \, d\ell \\
= 2t^3 + 6\pi t^2 \log(1 + Lt/3\pi),
\]
whence
\[
\hat{\varphi}^{[L]}(t) - \varphi^{[L]}(t) \leq \frac{C_1 e^{16\pi \sqrt{C_1\xi(3/2)}}}{(1 + \frac{Lt}{3\pi})^2} \left( (2 + r(C_1)^2)t^3 + 6\pi t^2 \log(1 + Lt/3\pi) \right) \\
\leq \frac{9\pi^2 C_1 e^{16\pi \sqrt{C_1\xi(3/2)}}}{L^2} \left( (2 + r(C_1)^2)t + 6\pi \log(1 + Lt/3\pi) \right) \\
\leq \frac{54\pi^3 C_1(2 + r(C_1)^2)e^{16\pi \sqrt{C_1\xi(3/2)}} \log(1 + L)}{L^2}.
\]
In the final line, we are simply shuffling constants using $t \leq \pi$. This completes the proof of (F.2).

**Derived estimates.** The remaining claims can be derived from the main claim we have just established; we will do so now. Below, we write $C_0 = \frac{54\pi^3 C_1(2 + r(C_1)^2)e^{16\pi \sqrt{C_1\xi(3/2)}}}{L^2}$. We will also assume $\ell \geq 1$.

We prove the claim about $\xi_\ell$ first. First, notice that for nonnegative numbers $a$, $b$, one has $1 - a - b \leq e^{2b}(1 - a)$ provided $a \leq 1/2$. Since $\varphi \leq \pi/2$, we have for each $\ell > 0$
\[
\xi_\ell(t) \leq \prod_{\ell' = \ell}^{L-1} \left( 1 - \frac{\hat{\varphi}^{[\ell']} (t)}{\pi} + \frac{C_0 \log(1 + \ell')}{\pi (\ell')^2} \right) \\
\leq \exp \left( 2C_0 \sum_{\ell' = \ell}^{L-1} \frac{\log(1 + \ell')}{(\ell')^2} \right) \xi_\ell(t).
\]
By the integral test estimate, we have for $\ell > 0$
\[
\sum_{\ell' = \ell}^{L-1} \frac{\log(1 + \ell')}{(\ell')^2} \leq \frac{\log(1 + \ell)}{\ell^2} + \int_\ell^L \frac{\log(1 + \ell')}{(\ell')^2} \, d\ell' \\
\leq \frac{\log(1 + \ell)}{\ell^2} + \log \left( 1 + \frac{1}{1 + \frac{1}{\ell}} \right) + \log(1 + \ell)/\ell - \log(1 + L)/L \\
\leq \frac{3\log(1 + \ell)}{\ell},
\]
where we applied $\log(1 + x) \leq x$ for all $x > -1$, whence for $\ell > 0$
\[
\xi_\ell(t) \leq e^{6C_0 \log(1 + \ell)/\ell} \xi_\ell(t).
\]
In particular, using the fact that $\log(1 + \ell)/\ell \leq 1$ and the estimate $e^{\epsilon x} \leq 1 + xe^\epsilon$ for $x \in [0, 1]$ (by convexity of the exponential function), we obtain
\[
\xi_\ell(t) \leq \left( 1 + e^{6C_0 \log(1 + \ell)/\ell} \right) \xi_\ell(t),
\]
as claimed. The proof of the second inequality is very similar: first, repeated application of the chain rule gives
\[
\hat{\varphi}^{[L]} = \prod_{\ell=0}^{L-1} \varphi \circ \varphi^{[\ell]}.
\]
Using the expression
\[
\hat{\varphi}(t) = \frac{(1 - t/\pi) \sin t}{\sin \varphi(t)},
\]

64
we can exploit a telescopic cancellation in the preceding expression for \( \hat{\phi}^{[L]} \), obtaining

\[
\hat{\phi}^{[L]} = \frac{\sin t}{\sin \hat{\phi}^{[L]}(t)} \prod_{\ell=0}^{L-1} \left(1 - \frac{\phi^{[\ell]}(t)}{\pi}\right).
\]

As the form of this upper bound is identical to the one we controlled for \( \xi_\ell \), only with a different constant factor, we can now apply the first part of that argument to the present setting, obtaining a bound

\[
\hat{\phi}^{(L)} \leq \frac{\sin t}{\sin \hat{\phi}^{[L]}(t)} \exp \left(\sum_{\ell=1}^{L-1} (\hat{\phi}^{[\ell]} - \phi^{[\ell]})\right) \exp \left(-\frac{1}{\pi} \sum_{\ell=0}^{L-1} \frac{t}{1 + \ell t/(3\pi)}\right)
\]

where in simplifying we also used that \( \phi^{[0]} = \hat{\phi}^{[0]} \). To proceed, we split the first sum, obtaining for any index \( 1 \leq \ell_* \leq L - 1 \)

\[
\sum_{\ell=1}^{L-1} (\hat{\phi}^{[\ell]} - \phi^{[\ell]}) = \sum_{\ell=1}^{\ell_*-1} (\hat{\phi}^{[\ell]} - \phi^{[\ell]}) + \sum_{\ell=\ell_*}^{L-1} (\hat{\phi}^{[\ell]} - \phi^{[\ell]})
\]

\[
\leq C t \sum_{\ell=1}^{\ell_*-1} 1 + 3C_0 \frac{\log(1 + \ell_*)}{\ell_*}
\]

\[
\leq C t \log(\ell_*) + 3C_0 \frac{\log(1 + \ell_*)}{\ell_*}
\]

\[
\leq C \log(1 + \ell_*) \left(1 + \frac{1}{\ell_*}\right),
\]

where in the second line the bound on the first sum used \( (F.5) \), and the second used the estimate we proved in the previous section and the integral test estimate above; in the third line we estimated the harmonic series with the integral test; and in the fourth line we worst-cased. Next, for any \( t \leq 1/\ell_* \), we have by the above

\[
\sum_{\ell=1}^{L-1} \hat{\phi}^{[\ell]} - \phi^{[\ell]} \leq C \log(1 + \ell_*)/\ell_*,
\]

and because the RHS approaches 0 as \( \ell_* \to \infty \), for any \( 0 < \varepsilon \leq 1 \) there is an integer \( N(\varepsilon) > 0 \) such that for all \( \ell_* \geq N \) we have

\[
\sum_{\ell=1}^{L-1} \hat{\phi}^{[\ell]} - \phi^{[\ell]} \leq \log(1 + \varepsilon).
\]

In particular, obtaining a lower bound for the RHS by concavity of \( \log \), it is sufficient to take \( \ell_* \geq C \varepsilon^{-2} \) for a suitably large absolute constant \( C > 0 \). To ensure there exists such a value of \( \ell_* \), it suffices to choose \( L \geq C \varepsilon^{-2} \) and therefore \( t \leq C' \varepsilon^2 \). In particular, plugging this estimate into our previous bound, we have shown that for any \( \varepsilon > 0 \), if \( L \geq C' \varepsilon^2 \) then for all \( t \leq C \varepsilon^2 \) we have

\[
\hat{\phi}^{[L]} \leq (1 + \varepsilon) \frac{\sin t}{\sin \hat{\phi}^{[L]}(t)} \exp \left(-\frac{1}{\pi} \sum_{\ell=0}^{L-1} \frac{t}{1 + \ell t/(3\pi)}\right).
\]

We then calculate by an estimate from the integral test

\[
\sum_{\ell=0}^{L-1} \frac{t}{1 + \ell t/(3\pi)} \geq \int_0^L \frac{t}{1 + \ell t/(3\pi)} \, d\ell = 3\pi \log(1 + Lt/(3\pi)),
\]

which establishes under the previous conditions on \( L \) and \( t \) that

\[
\hat{\phi}^{[L]} \leq \frac{\sin t}{\sin \hat{\phi}^{[L]}(t)} \frac{1 + \varepsilon}{(1 + Lt/(3\pi))^2}.
\]

To conclude, we need to simplify the \( \sin \) ratio term. Using Lemma \( F.4 \), for any \( 0 < \varepsilon' \leq 1/2 \), we have for \( 0 \leq t \leq C \varepsilon' \) that

\[
\frac{\sin t}{\sin \hat{\phi}^{[L]}(t)} \leq (1 + 2\varepsilon')(1 + Lt/(3\pi)),
\]

65
which suffices to prove the claim for small \( t \) after noting \( (1 + 2\varepsilon')(1 + \varepsilon) = 1 + 2\varepsilon' + \varepsilon + 2\varepsilon'\varepsilon \), choosing whichever is smaller, and adjusting the preceding conditions on \( t \) and \( L \) (i.e. the absolute constants in the previous bounds may grow/shrink as necessary). To show the claimed bound on the entire interval \([0, \pi]\), we can follow exactly the argument above, but instead of partitioning the sum of errors \( \hat{\varphi}^{[\ell]} - \varphi^{[\ell]} \) as above we simply use bound the sum of errors as in the bound on \( \hat{\xi}_\ell \) previously to obtain a large constant in the numerator; the sin ratio is controlled in this case using the first conclusion in Lemma F.4, which is valid on the whole interval \([0, \pi]\).

Finally, we obtain the estimate on \( \psi \) by calculating using the estimate involving \( \xi_\ell \) and \( \hat{\xi}_\ell \) that we proved earlier. First, we note that although we required \( \ell > 0 \) above, the fact that \( \hat{\varphi}^{[0]} = \varphi^{[0]} \) implies that we have an estimate \( \xi_0 \leq (1 + \log(2)e^{6C_0})\hat{\xi}_0 \). We therefore have

\[
\psi(t) = \frac{n}{2} \sum_{\ell=0}^{L-1} \xi_\ell(t) \leq \frac{n}{2} \sum_{\ell=0}^{L-1} \left( 1 + e^{6C_0} \log \left( \frac{1 + \ell}{\ell} \right) \right) \hat{\xi}_\ell(t)
\leq \hat{\psi}(t) + \left( \frac{n}{2} \right) e^{6C_0} \left( \log(2)\hat{\xi}_0(t) + \sum_{\ell=1}^{L-1} \frac{\log(1 + \ell)}{\ell} \hat{\xi}_\ell(t) \right).
\]

It is easy to see that \( \hat{\xi}_\ell \leq 1 \). Hence

\[
\psi(t) \leq \hat{\psi}(t) + \left( \frac{n}{2} \right) e^{6C_0} \left( \log(2) + \sum_{\ell=1}^{L-1} \frac{\log(1 + \ell)}{\ell} \right)
\leq \hat{\psi}(t) + ne^{6C_0} \left( \log(2) + \sum_{\ell=2}^{L-1} \frac{\log(\ell)}{\ell} \right)
\leq \hat{\psi}(t) + ne^{6C_0} \left( 2 \log(2) + \int_{\ell=2}^{L-1} \frac{\log(\ell)}{\ell} \, d\ell \right)
\leq \hat{\psi}(t) + ne^{6C_0} \left( 2 \log(2) - (1/2) \log^2 2 + (1/2) \log^2 (L - 1) \right)
\leq \hat{\psi}(t) + 4ne^{6C_0} \log^2 L,
\]

where the final bound requires \( L \geq 3 \).

\[\blacksquare\]

**Lemma F.4.** For \( \ell \in \mathbb{N}_0 \), one has for \( t \in [0, \pi] \)

\[
\frac{\sin(t)}{\sin(\varphi^{[\ell]}(t))} \leq 3(1 + \ell t/(3\pi))
\]

and there exists an absolute constant \( C > 0 \) such that for any \( 0 < \varepsilon \leq 1/2 \), if \( 0 \leq t \leq C\varepsilon \) one has

\[
\frac{\sin(t)}{\sin(\varphi^{[\ell]}(t))} \leq (1 + 2\varepsilon)(1 + \ell t/(3\pi)).
\]

**Proof.** We prove the bound on \([0, \pi]\) first. Because \( t \mapsto t^{-1} \sin t \) is decreasing on \([0, \pi]\), we apply Lemma G.6 to get

\[
\frac{\sin(t)}{\sin(\varphi^{[\ell]}(t))} \leq \frac{t}{\varphi^{[\ell]}(t)}
\leq \frac{t}{1 + \ell t/(\pi)}
\leq 1 + \ell t/(3\pi) \leq 3(1 + \ell t/(3\pi)).
\]

Now fix \( 0 < \varepsilon \leq 1/2 \). We claim that there is an absolute constant \( C > 0 \) such that if \( t \leq C\varepsilon \), we have

\[
\varphi^{[\ell]}(t) \geq (1 - \varepsilon)\varphi^{[\ell]}(t).
\]

Thus our claim implies the desired bound.

66
Assuming this claim, we have for \( t \leq C\epsilon \)

\[
\frac{\sin(t)}{\sin(\varphi^{[\ell]}(t))} \leq \frac{t}{\varphi^{[\ell]}(t)} \leq \frac{t}{(1 - \epsilon)\varphi^{[\ell]}(t)} \leq \frac{1}{1 - \epsilon} (1 + \ell t / (3\pi)) \leq (1 + 2\epsilon)(1 + \ell t / (3\pi)),
\]

which is enough to conclude after rescaling. Now we want to show the claim. Let \( C_0 = \max \{1, C_1\} \) where \( C_1 \) denotes the constant on \( \ell^3 \) in Lemma F.13. We first notice that

\[
\varphi(t) \geq \tilde{\varphi}(t) - C_0\ell^3 = \frac{t}{1 + \ell t / (3\pi)} - C_0\ell^3 = \frac{t}{1 + \ell t / (3\pi)} - \frac{t}{1 + \pi / (3\pi)} \frac{4}{3} C_0\ell^2 \geq (1 - \frac{4}{3} C_0\ell^2) \frac{t}{1 + \ell t / (3\pi)}.
\]

We are going to proceed with an induction-like approach. Put \( \epsilon_1 = 4C_0\ell^2 / 3 \), and choose \( t \leq \sqrt{3 / (4C_0)} \) so that \( 1 - \epsilon_1 \geq 0 \). Supposing that it holds \( \varphi^{[\ell-1]} \geq (1 - \epsilon_{\ell-1})\tilde{\varphi}^{[\ell-1]}(t) \) for a positive \( \epsilon_{\ell-1} \) such that \( 1 - \epsilon_{\ell-1} \geq 0 \) (we have shown there is such \( \epsilon_1 \) and controlled it), we have by some applications of the induction hypothesis, Lemma F.1, and the previous small-\( t \) estimate (we use below that \( t \leq \sqrt{3 / (4C_0)} \))

\[
\varphi^{[\ell]}(t) \geq \left(1 - \frac{4}{3} C_0 \left(\varphi^{[\ell-1]}(t)\right)^2\left(\varphi^{[\ell-1]}(t)\right)\right) \tilde{\varphi}(\varphi^{[\ell-1]}(t)) = \left(1 - \frac{4}{3} C_0 \left(\varphi^{[\ell-1]}(t)\right)^2\right) \frac{(1 - \epsilon_{\ell-1})\frac{t}{1 + (\ell - 1) t / (3\pi)}}{1 + (3\pi)^{-1} (1 - \epsilon_{\ell-1}) \frac{t}{1 + (\ell - 1) t / (3\pi)}} \geq (1 - \frac{4}{3} C_0 \left(\varphi^{[\ell-1]}(t)\right)^2) \frac{t^2}{1 + \ell t / (3\pi) - \epsilon_{\ell-1} t / (3\pi)} \geq (1 - \frac{4}{3} C_0 \left(\varphi^{[\ell-1]}(t)\right)^2) \frac{t^2}{1 + (\ell - 1) t / (3\pi)} \geq (1 - \frac{4}{3} C_0 \left(\varphi^{[\ell-1]}(t)\right)^2) \frac{t^2}{1 + (\ell - 1) t / (3\pi)} - \epsilon_{\ell-1} \tilde{\varphi}^{[\ell]}(t).
\]

This shows that we can take \( \epsilon_{\ell} = \epsilon_{\ell-1} + (4C_0 / 3) t^2 / (1 + (\ell - 1) t / (3\pi))^2 \) as long as this term is not larger than 1. Unraveling inductively to check, we get

\[
\epsilon_{\ell} = \sum_{\ell' = 0}^{\ell-1} \frac{4C_0}{3} \frac{t^2}{(1 + \ell' t / (3\pi))^2} \leq \frac{4C_0}{3} \frac{t^2}{3 \pi} \left(1 + \frac{\ell - 1}{1 + (\ell - 1) t / (3\pi)}\right) \leq \frac{4C_0}{3} \frac{t^2}{3 \pi} \left(1 + \frac{\ell - 1}{1 + (\ell - 1) t / (3\pi)}\right) \leq \frac{16\pi C_0}{3} t.
\]

In particular, the induction is consistent as long as \( t \leq 3 / (16\pi C_0) \). Note as well that since \( C_0 \geq 1 \) we have \( \sqrt{3 / (4C_0)} \geq 3 / (16\pi C_0) \). Thus by induction, we know that when \( 0 < \epsilon < 1 \) and \( t \leq \frac{3C_0\pi}{16\pi} \),
we have
\[ \varphi^{[\ell]}(t) \geq (1 - \epsilon)\hat{\varphi}^{[\ell]}(t) \]
as claimed. \hfill \square

### F.2 Sharp Lower Bound for \( \psi \)

**Lemma F.5.** There is an absolute constant \( C_0 > 0 \) such that
\[ \psi(\pi) \leq \frac{n(L - 1)}{8} + 6\pi ne^{6C_0} \log^2 L \]

**Proof.** Following Lemma F.3 (worsening constants slightly for convenience), we directly have
\[ \psi(\pi) \leq \hat{\psi}(\pi) + 6\pi ne^{6C_0} \log^2 L. \]

\( \hat{\psi}(t) \) has a closed form expression, by notice that
\[
\hat{\xi}_t(t) = \prod_{\ell' = t}^{L-1} \left( 1 - \frac{\hat{\varphi}^{[\ell']}(t)}{\pi} \right)
= \prod_{\ell' = t}^{L-1} \left( 1 - \frac{t/\pi}{1 + \ell't/(3\pi)} \right)
= \prod_{\ell' = t}^{L-1} \frac{1 + (\ell' - 3)t/(3\pi)}{1 + \ell't/(3\pi)}
= \frac{(1 + (\ell' - 3)t/(3\pi)) (1 + (\ell' - 2)t/(3\pi)) (1 + (\ell' - 1)t/(3\pi))}{(1 + (L - 3)t/(3\pi)) (1 + (L - 2)t/(3\pi)) (1 + (L - 1)t/(3\pi))}. \tag{F.8}
\]
and
\[
\hat{\psi}(t) = \frac{n}{2} \sum_{t=0}^{L-1} \hat{\xi}_t(t)
= \frac{n}{2} \sum_{t=0}^{L-1} \frac{(3\pi + (\ell - 3)t)(3\pi + (\ell - 2)t)(3\pi + (\ell - 1)t)}{(3\pi + (L - 3)t)(3\pi + (L - 2)t)(3\pi + (L - 1)t)}
= \frac{n}{2} \frac{1}{4t} \sum_{t=0}^{L-1} \frac{(3\pi + (\ell - 3)t)(3\pi + (\ell - 2)t)(3\pi + (\ell - 1)t)(3\pi + \ell t)}{(3\pi + (L - 3)t)(3\pi + (L - 2)t)(3\pi + (L - 1)t)}
- \sum_{t=0}^{L-1} \frac{(3\pi + (\ell - 3)t)(3\pi + (\ell - 2)t)(3\pi + (\ell - 1)t)}{(3\pi + (L - 3)t)(3\pi + (L - 2)t)(3\pi + (L - 1)t)}
= \frac{n}{8t} \frac{(3\pi + (L - 4)t)(3\pi + (L - 3)t)(3\pi + (L - 2)t)(3\pi + (L - 1)t)}{(3\pi + (3\pi - 4t)(3\pi - 3t)(3\pi - 2t)(3\pi - t)}
- \frac{(3\pi + (L - 3)t)(3\pi + (L - 2)t)(3\pi + (L - 1)t)}{(3\pi + (3\pi - 4t)(3\pi - 3t)(3\pi - 2t)(3\pi - t)}
= \frac{n}{8t} \frac{(3\pi + (L - 4)t)(3\pi + (L - 3)t)(3\pi + (L - 2)t)(3\pi + (L - 1)t)}{(3\pi + (3\pi - 4t)(3\pi - 3t)(3\pi - 2t)(3\pi - t)}
= \frac{n(L - 4)}{8} + \frac{n}{8t} \frac{(3\pi - (3\pi - 4t)(3\pi - 3t)(3\pi - 2t)(3\pi - t)}{(3\pi + (3\pi - 4t)(3\pi - 3t)(3\pi - 2t)(3\pi - t)} \tag{F.9}
\]
From the second to the fourth line above, we used a telescopic series cancellation trick to sum. Then we get the claim as
\[
\psi(\pi) \leq \hat{\psi}(\pi) + 6\pi ne^{6C_0} \log^2 L
= \frac{n(L - 4)}{8} + \frac{3\pi n}{8t} + 6\pi ne^{6C_0} \log^2 L
\]
As a result, we have

\[ \frac{n(L-1)}{8} + 6\pi n e^{6c_0} \log^2 L. \]

\[ \square \]

**Lemma F.6.** When \( L \geq 2 \), we have for any \( r > 0 \)

\[
\int_0^r \psi(t)dt \geq \frac{n(L-4)}{8} + \frac{3\pi n}{8} \log(1 + \frac{L-2}{3\pi r})
\]

**Proof.** From Lemma F.2, we have \( \psi^{[\ell]}(t) \leq \frac{t^\ell}{1 + t^\ell/(3\pi)} \). Thus we get

\[
\xi(t) = \prod_{\ell'=0}^{L-1} (1 - \frac{1}{\pi} \phi^{[\ell']}(t))
\]

\[
\geq \frac{(3\pi + (\ell-3)t)(3\pi + (\ell-2)t)(3\pi + (\ell-1)t)}{(3\pi + (L-3)t)(3\pi + (L-2)t)(3\pi + (L-1)t)}.
\]

As a result, we have

\[
\psi(t) = \frac{n}{2} \sum_{\ell=0}^{L-1} \xi(t)
\]

\[
\geq \frac{n}{2} \sum_{\ell=0}^{L-1} \frac{(3\pi + (\ell-3)t)(3\pi + (\ell-2)t)(3\pi + (\ell-1)t)}{(3\pi + (L-3)t)(3\pi + (L-2)t)(3\pi + (L-1)t)}
\]

\[
= \frac{n}{2} \sum_{\ell=0}^{L-1} \frac{(3\pi + (\ell-4)t)(3\pi + (\ell-3)t)(3\pi + (\ell-2)t)(3\pi + (\ell-1)t)}{(3\pi + (L-3)t)(3\pi + (L-2)t)(3\pi + (L-1)t)}
\]

\[
= \frac{n}{8t} \frac{(3\pi + (L-3)t)(3\pi + (L-2)t)(3\pi + (L-1)t)}{(3\pi + (L-4)t)(3\pi + (L-3)t)(3\pi + (L-2)t)(3\pi + (L-1)t)}
\]

\[
\geq \frac{n}{8t} \frac{(3\pi + (L-3)t)(3\pi + (L-2)t)(3\pi + (L-1)t)}{(3\pi + (L-4)t)(3\pi + (L-3)t)(3\pi + (L-2)t)(3\pi + (L-1)t)}
\]

\[
= \frac{n}{8t} \frac{(3\pi - 4t)(3\pi - 3t)(3\pi - 2t)(3\pi - t)}{(3\pi + (L-4)t)(3\pi + (L-3)t)(3\pi + (L-2)t)(3\pi + (L-1)t)}
\]

\[
= \frac{n}{8t} \left( 1 - \frac{1}{(1 + (L-3)t)(1 + (L-2)t)(1 + (L-1)t)} \right)
\]

\[
\geq \frac{n}{8t} \left( 1 - \frac{1}{(1 + (L-2)t)^3} \right)
\]

\[
= \frac{n}{8t} \left( 1 + \frac{L'}{1 + L't} \right)
\]

\[
= \frac{n(L-4)}{8} + \frac{n}{8t} \left( \frac{L'}{1 + L't} + \frac{L'}{(1 + L't)^2} + \frac{L'}{(1 + L't)^3} \right).
\]

In the third and fourth lines above, we used a splitting and cancellation trick to sum similar to what we used in Lemma F.5. In moving from the seventh to the eighth line, we used the inequality \((x-1)(x+1) \leq x^2\) after splitting off a term that can be negative for large \( t' \). In moving from
the eighth to the ninth line, we used nonnegativity of the third summand and upper bounded the numerator of the term in the second summand. (In both of the previous simplifications, we are using that \( t' \leq 1/3 \).) The remaining simplifications obtain a common denominator in the second term and then cancel. Integrating, we thus find

\[
\int_0^r \psi(t)dt \geq \frac{n(L - 4)}{8} r + \frac{3\pi n}{8} \left( \log(1 + L' \frac{r}{3\pi}) + \left(1 - \frac{1}{1 + L'r / 3\pi}\right) + \frac{1}{2} \left(1 - \frac{1}{(1 + L'r / 3\pi)^2}\right) \right)
\]

when \( L' \geq 0 \) \( n(L - 4) \frac{r}{8} + \frac{3\pi n}{8} \log(1 + L - 2 \frac{r}{3\pi}) \).

\[ \blacksquare \]

**Lemma F.7.** There exists an absolute constant \( C > 0 \) such that when \( L \geq 2 \), we have for any \( r > 0 \)

\[
\int_0^r (\psi(t) - \psi(\pi))dt \geq \frac{3\pi n}{8} \log(1 + \frac{L - 2}{3\pi} r) - Cnr \log^2 L
\]

**Proof.** Following Lemma F.6 and Lemma F.5, we directly get

\[
\int_0^r (\psi(t) - \psi(\pi))dt \geq \int_0^r \psi(t)dt - \left( \frac{n(L - 1)}{8} + 6\pi n e^{6C_0} \log^2 L \right) r
\]

\[
\geq \frac{3\pi n}{8} \log(1 + \frac{L - 2}{3\pi} r) - 6\pi n e^{6C_0} \log^2 L r - \frac{3n}{8} r
\]

\[
\geq \frac{3\pi n}{8} \log(1 + \frac{L - 2}{3\pi} r) - \left( 6\pi e^{6C_0} + \frac{3}{8} \right) nr \log^2 L
\]

\[ \blacksquare \]

**F.3 Nearly-Matching Upper Bound**

**Lemma F.8.** There exist absolute constants \( C, C' > 0 \) and absolute constants \( K, K' > 0 \) such that for any \( 0 < \varepsilon \leq 1 \), if \( L \geq K\varepsilon^{-3} \) then for any \( 0 \leq t \leq K'\varepsilon^3 \) one has

\[
\psi(t) - \psi(\pi) \leq (1 + \varepsilon) \left( 1 + \frac{18}{1 + (L - 3)t/(3\pi)} + \frac{C \log^2(L)}{L} \right) \frac{n}{8} \frac{L - 3}{1 + (L - 3)t/(3\pi)}
\]

and for any \( 0 \leq t \leq \pi \) one has

\[
\psi(t) - \psi(\pi) \leq C'n \frac{L - 3}{1 + (L - 3)t/(3\pi)}.
\]

**Proof.** We try to control the DC subtracted skeleton \( \psi(t) - \psi(\pi) \) by its derivative \( \dot{\psi}(t) \), which would require us to control the derivatives \( \xi(t) \) and further \( \dot{\psi}[\ell](t) \). Fix \( 0 < \varepsilon \leq 1/2 \). When \( L \geq C_0 \varepsilon^{-2} \) for some constant \( C_0 > 0 \), Lemma F.3 provides sharp bound for \( \dot{\psi}[\ell](t) \) with

\[
\dot{\psi}[\ell](t) \leq \frac{1 + \varepsilon}{(1 + c\ell t)^2} \quad t \in [0, C'\varepsilon^2]
\]

\[
\dot{\psi}[\ell](t) \leq \frac{C_1}{(1 + c\ell t)^2} \quad t \in [0, \pi]
\]

with absolute constants \( C', C_1 > 0 \) and \( c = 1/(3\pi) \). For notation convenience, define \( t_1 = C'\varepsilon^2 \) and write

\[
M_t = \begin{cases} 
1 + \varepsilon, & 0 \leq t \leq t_1 \\
C_1, & \text{otherwise}
\end{cases}
\]

We can compactly write the previous two bounds together as

\[
\dot{\psi}[\ell](t) \leq \frac{M_t}{(1 + c\ell t)^2}.
\]

70
This allows us to separate \( \psi(t) - \psi(\pi) \) into two components \( \psi(t) - \psi(t_1) \) and \( \psi(t_1) - \psi(\pi) \), where we get the correct constant \( 1 + \epsilon \) in the first component and control the second component by the fact that \( \psi \) becomes sharp when \( L \) is large, making the difference between \( \psi(t_1) \) and \( \psi(\pi) \) negligible.

Now, for \( \ell \geq 4 \), with \( c = \frac{1}{3\pi} \), we have

\[
|\xi_\ell(t)| = \frac{\xi_\ell(t)}{\pi} \sum_{t'=\ell}^{L-1} \frac{\varphi(t')}{1 - \varphi(t')/\pi} \\
\leq \frac{\xi_\ell(t)}{\pi} \sum_{t'=\ell}^{L-1} \frac{\varphi(t')}{1 - \frac{t'/\pi}{1+c't}} \\
= \frac{\xi_\ell(t)}{\pi} \sum_{t'=\ell}^{L-1} \frac{1+c't}{1+c'(t'-3)t} \varphi(t') \\
\leq \frac{\xi_\ell(t)}{\pi} \sum_{t'=\ell}^{L-1} \frac{1+c't}{1+c'(t'-3)t} \frac{M_t}{(1+c'\ell t)^2}
\]

Let \( c_{1,\ell} = 1 + e^{6c_0 \log(1+\ell)/\ell} \). (F.3) and (F.8) provide control for \( \xi_\ell(t) \) and we have

\[
|\xi_\ell(t)| \leq \frac{c_{1,\ell}}{\pi} \frac{(1+c(\ell-3)t)(1+c(\ell-2)t)(1+c(\ell-1)t)}{(1+c(L-3)t)(1+c(L-2)t)(1+c(L-1)t)} \sum_{t'=\ell}^{L-1} \frac{1+c't}{1+c'(t'-3)t} \frac{M_t}{(1+c'\ell t)^2} \\
\leq \frac{c_{1,\ell}M_t}{\pi} \frac{(1+c(\ell-3)t)(1+c(\ell-2)t)(1+c(\ell-1)t)}{(1+c(L-3)t)(1+c(L-2)t)(1+c(L-1)t)} \int_{t'=\ell}^{L-1} \frac{1}{(1+c(t'-2)t)^2} dt' \\
\leq \frac{c_{1,\ell}M_t}{\pi} \frac{(1+c(\ell-3)t)(1+c(\ell-2)t)(1+c(\ell-1)t)}{(1+c(L-3)t)(1+c(L-2)t)(1+c(L-1)t)} \frac{L - \ell}{(1+c(\ell-3)t)^2(1+c(L-3)t)^2(1+c(L-2)t)(1+c(L-1)t)} \\
+ \frac{C \log(1+\ell)/\ell}{\ell} \frac{L}{(1+c(\ell-3)t)^2}.
\]

In moving from the fifth to the sixth line, we used that \( (1+c(\ell-3)t)(1+c(\ell)t) = 1 + c(2\ell - 3)t + c^2(\ell - 3)t^2 \geq 1 + c(2\ell - 4)t + c^2(\ell^2 - 4\ell + 4)t^2 = (1+c(\ell-2)t)^2 \) provided \( \ell \geq 4 \) and subsequently the integral test. In the splitting in the last line, we used that \( M_t \) is always bounded by a (very large) absolute constant, and worst-cased (as this term will be sub-leading in \( L \)).

To control derivatives of \( \psi \), we need to control sums of the derivatives above. We will derive some further estimates for this purpose. First, we calculate

\[
\sum_{t=1}^{L-1} \frac{(1+c(\ell-2)t)(1+c(\ell-1)t)}{(1+c(L-2)t)(1+c(L-1)t)}(L - \ell) \\
= \sum_{t=1}^{L-1} \left( (L - \ell) + (L - \ell)(2\ell - 3)ct + (L - \ell)(\ell - 1)(\ell - 2)c^2t^2 \right) \\
= \frac{L(L-1)}{2} + \frac{L(L-1)(L-\frac{3}{2})}{3}ct + \frac{L(L-1)(L-2)(L-3)}{12}c^2t^2 \\
\leq \frac{L(L-1)}{2} + \frac{L(L-1)(L-\frac{3}{2})}{3}ct + \frac{L(L-3)}{12} + 1 \frac{L(L-3)}{12} \frac{1}{(1+c(L-1)t)(1+c(L-2)t)} \\
\leq \frac{6L(L-1)}{L-3} + 4ct(L-1) + 1 \frac{L(L-3)}{12} \frac{1}{(1+c(L-1)t)(1+c(L-2)t)}.
\]

If \( L \geq 4 \), we can simplify a term in the last line of the previous expression as

\[
\frac{6L(L-1)}{L-3} + 4ct(L-1) \frac{18}{(1+c(L-1)t)(1+c(L-2)t)} \leq \frac{18}{1+c(L-1)t},
\]
and under \( L \geq 4 \) we also have
\[
\sum_{\ell=2}^{L-1} \log(1 + \frac{1}{\ell}) \lesssim \log(L) \int_{1}^{L-1} \frac{1}{t} \lesssim \log^2 L.
\]

Applying the upper bound from before and adding some terms to the sum (because all terms are nonnegative), we get
\[
\pi \sum_{\ell=4}^{L-1} \frac{\xi_{\ell}(t)}{L} \leq \left( 1 + \frac{18}{1 + c(L-1)t} \right) \frac{L(L - 3)}{12} \frac{M_t}{(1 + c(L-3)t)^2} + \frac{CL \log^2 L}{(1 + c(L-3)t)^2}. \tag*{(F.10)}
\]

From Lemma G.9, for \( \ell = \{0, 1, 2, 3\} \), we can bound \( \xi_{\ell}(t) \leq \frac{1+Lt/\pi}{1+Lt/\pi} \). Using that \( \xi_{\ell} \) is decreasing for all \( \ell \geq 0 \) and nonnegative, for \( t, t' \in [0, \pi], t' \geq t \), we are now able to control the DC subtracted skeleton as
\[
\psi(t) - \psi(t') \\
\leq \frac{n}{2} \left( \frac{4 + (0 + 1 + 2 + 3)t/\pi}{1 + Lt/\pi} \right) \sum_{\ell=4}^{L-1} \int_{t=t}^{t'} \frac{\xi_{\ell}(v)}{v} dv
\leq \frac{2 + 3t/\pi}{1 + Lt/\pi} n + \frac{n}{2\pi} \int_{t=t}^{t'} \left( \frac{1 + \frac{18}{1 + c(L-1)v}}{1 + c(L-3)v} \right) \frac{M_c(L-3)/12}{(1 + c(L-3)v)^2} + \frac{C \log^2 L}{(1 + c(L-3)v)^2} dv
\leq \frac{2 + 3t/\pi}{1 + Lt/\pi} n + \frac{n}{2\pi} \int_{t=t}^{t'} \left( \frac{1 + \frac{18}{1 + c(L-1)v}}{1 + c(L-3)v} \right) \frac{M_c(L-3)/12 + C \log^2 L}{(1 + c(L-3)v)^2} dv
\leq \frac{2 + 3t/\pi}{1 + Lt/\pi} n + \frac{n}{2\pi} L \left( \frac{1 + \frac{18}{1 + c(L-1)t}}{12} + C \log^2 L \right) \int_{v=t}^{\pi} \frac{1}{(1 + c(L-3)v)^2} dv.
\]

From the second to the third line, we use the fact that \( M_c \) is nondecreasing in \( v \). Thus we have
\[
\psi(t) - \psi(t') \leq \frac{2 + 3t/\pi}{1 + Lt/\pi} n + \frac{n}{2\pi} L \left( \frac{1 + \frac{18}{1 + c(L-1)t}}{12} + C \log^2 L \right) \nu \bigg|_{v=t}^{\pi}
\leq \frac{5n}{1 + Lt/\pi} + \frac{n}{2} \left( \frac{L}{1 + c(L-3)\pi} \right) \frac{M_c(L-3)}{24} \left( \frac{1 + \frac{18}{1 + c(L-1)t}}{1 + c(L-3)t} \right) + \frac{12C \log^2 L}{L-3}
\leq \frac{5n}{1 + Lt/\pi} + \frac{L-3}{1 + c(L-3)\pi} \frac{M_c}{24} \left( \frac{1 + \frac{18}{1 + c(L-1)t}}{1 + c(L-3)t} \right) + \frac{12C \log^2 L}{L-3}
\leq \frac{5n}{1 + Lt/\pi} + \frac{L-3}{1 + c(L-3)\pi} \frac{M_c}{24} \left( \frac{1 + \frac{18}{1 + c(L-1)t}}{1 + c(L-3)t} \right) + \frac{12C \log^2 L}{L-3}
\leq \frac{nM_c}{8} \left( \frac{L-3}{1 + c(L-3)\pi} \right) \left( \frac{1 + \frac{18}{1 + c(L-1)t}}{1 + c(L-3)t} \right) + \frac{C \log^2 L}{L-3}
\leq \frac{nM_c}{8} \left( \frac{L-3}{1 + c(L-3)\pi} \right) \left( \frac{1 + \frac{18}{1 + c(L-1)t}}{1 + c(L-3)t} \right) + \frac{C \log^2 L}{L}.
\]

In moving from the second to the third line, we simplified/rearranged and used that \( C_1 \geq 1 \). In moving from the fourth to the fifth line, we replace the numerator of \( L \) in the leading term with \( L - 3 + 3 \), then expand and simplify. In particular we have
\[
\psi(t_1) - \psi(\pi) \leq \frac{nC_1}{8} \frac{1}{\alpha t_1} \left( 19 + \frac{C \log^2 L}{L} \right)
\]
And if would give us the claim.

This inequality holds for all \( t \) because when \( t \geq t_1 \), the left hand side is negative. Notice that when \( t \leq \epsilon^3/(2C_1C'(19+C)) \) and \( L - 3 \geq C_1C'(19+C)/(\epsilon^3) \), we would have

\[
\frac{C_1C'(19+C)}{8\epsilon^2} \leq L - 3
\]

Combining the results and notice that we can absorb the factor of \( \epsilon \) and \( r \), specifically set \( r = \min\{b/t_1\} \). From Lemma F.8 and (F.11), we would have

\[
\int_a^b (\psi(t) - \psi(\pi))dt \leq Cn \log \left( \frac{1 + (L - 3)b/(3\pi)}{1 + (L - 3)a/(3\pi)} \right).
\]

(F.11)

And if \( r > 0 \) satisfies \( r \leq K' \epsilon^3 \), one further has

\[
\int_r^b (\psi(t) - \psi(\pi))dt \leq (1 + \varepsilon)^3 \frac{3\pi n}{8} \log \left( \frac{1 + (L - 3)b/(3\pi)}{1 + (L - 3)r/(3\pi)} \right) + C'n \log(\pi/(K'\epsilon^3)).
\]

(F.12)

Proof. (F.11) follows directly from Lemma F.8 and integration. To achieve an upper bound for integral from \( r \) to \( b \), we cut the integral at \( t_1 = K' \epsilon^3 \) and apply bounds from Lemma F.8 separately. Specifically, set \( b' = \min\{b, t_1\} \), from Lemma F.8 and (F.11) we would have

\[
\int_r^b (\psi(t) - \psi(\pi))dt = \int_r^{b'} (\psi(t) - \psi(\pi))dt + \int_{b'}^b (\psi(t) - \psi(\pi))dt
\]

\[
\leq (1 + \varepsilon) \left( 1 + \frac{C\log^2(L)}{L} \right) \frac{3\pi n}{8} \log \left( \frac{1 + (L - 3)b/(3\pi)}{1 + (L - 3)r/(3\pi)} \right) + \frac{18n(1 + \varepsilon)}{8} \frac{L - 3}{(1 + (L - 3)t/(3\pi))^2}dt + C'n \log \left( \frac{1 + (L - 3)b/(3\pi)}{1 + (L - 3)\pi/(3\pi)} \right)
\]

\[
\leq (1 + \varepsilon) \left( 1 + \frac{C\log^2(L)}{L} \right) \frac{3\pi n}{8} \log \left( \frac{1 + (L - 3)b/(3\pi)}{1 + (L - 3)r/(3\pi)} \right) + \frac{9n}{2(3\pi)^2} + C'n \log(b/b')
\]

\[
\leq (1 + \varepsilon) \frac{3\pi n}{8} \log \left( \frac{1 + (L - 3)b/(3\pi)}{1 + (L - 3)r/(3\pi)} \right) + 2 \frac{C\log^2(L)}{L} \frac{3\pi n}{8} \log(1 + (L - 3)\pi/(3\pi)) + \frac{n}{2\pi^2} + C'n \log(\pi/t_1)
\]

\[
\leq (1 + \varepsilon) \frac{3\pi n}{8} \log \left( \frac{1 + (L - 3)b/(3\pi)}{1 + (L - 3)r/(3\pi)} \right) + \frac{C\log^3(L)}{L} + C'n \log(\pi/(K'\epsilon^3)).
\]

(F.12) then follows by setting \( L \geq K \epsilon^{-3} \) for some \( K > 0 \).
F.4 Higher Order Derivatives of $\psi$

Lemma F.10. There exist absolute constants $C, C'$ such that when $L \geq C$, we have for any $r \in [0, \pi]$, 
\[
\max_{t \geq r} |\psi(t)| \leq \frac{C'n}{r^2} \tag{F.13}
\]
and we can control the integration
\[
\int_r^t t^3 |\dot{\psi}(t)| \, dt \leq C'nr^2 \tag{F.14}
\]

Proof. From (F.10) (we control $M_t \leq C$ for an absolute constant $C > 0$ in this context, so that we do not need to deal with the conditions on $\varepsilon$ that appear there) and Lemma G.10, we have
\[
|\dot{\psi}(t)| \leq \frac{n}{2} \sum_{\ell=0}^{3} |\dot{\xi}_\ell(t)| + \frac{Cn}{2} \left( \left( 1 + \frac{18}{1 + (L - 1)t/(3\pi)} \right) \frac{L(L-3)}{12} \frac{1}{(1 + (L - 3)t/(3\pi))^2} + \frac{L \log^2 L}{1 + (L - 3)t/(3\pi))^2} \right) 
\]
\[
\leq \frac{6\pi n}{t} + \frac{Cn}{2} \left( \frac{L(L-3)}{(L-3)^2t^2} + \frac{L \log^2 L}{(L-3)^2t^2} \right) 
\]
\[
\leq \frac{Cn}{t^2}.
\]
This directly get us (F.13) and (F.14). \qed

Lemma F.11. There exist absolute constants $C, C'$ such that when $L \geq C$, we have for any $r > 0$
\[
\max_{t \geq r} |\ddot{\psi}(t)| \leq \frac{C'r n}{r^3} \tag{F.15}
\]
and
\[
\int_0^\pi t^6 |\ddot{\psi}(t)| \, dt \leq C'n.
\]

Proof. Following Lemmas G.8, G.10 and G.11, we have
\[
\dot{\xi}_\ell = -\frac{\xi_1}{\pi} I_{t=0} - \frac{\xi_\ell}{\pi} \sum_{\ell'=\max\{1, \ell\}}^{L-1} \frac{\dot{\varphi}[\ell']}{1 - \varphi[\ell']/\pi} \tag{F.16}
\]
and
\[
\dot{\xi}_\ell = -\frac{\xi_\ell}{\pi} \sum_{\ell'=\max\{1, \ell\}}^{L-1} \frac{\dot{\varphi}[\ell']}{1 - \varphi[\ell']/\pi} + \frac{\xi_\ell}{\pi^2} \sum_{\ell', \ell'':\max\{1, \ell\}}^{L-1} \frac{\dot{\varphi}[\ell']\dot{\varphi}[\ell'']}{(1 - \varphi[\ell']/\pi)(1 - \varphi[\ell'']/\pi)}
\]
\[
- \frac{2\xi_\ell}{\pi^2} \sum_{\ell'=1}^{L-1} \frac{\dot{\varphi}[\ell']}{1 - \varphi[\ell']/\pi}
\]
\[
\leq \left| \frac{\xi_\ell}{\pi} \sum_{\ell'=\max\{1, \ell\}}^{L-1} \frac{\dot{\varphi}[\ell']}{1 - \varphi[\ell']/\pi} \right|.
\]
which gives
\[ \xi \frac{L}{\pi^2} \left( \sum_{\ell' = \max(1, \ell)}^{L-1} \frac{\phi^{(\ell')}}{1 - \phi^{(\ell')}/\pi} \right)^2 + \frac{2 \xi_1}{\pi^2} \sum_{\ell' = 0}^{L-1} \frac{\phi^{(\ell')}}{1 - \phi^{(\ell')}/\pi}, \]
where the diagonal is added to obtain the upper bound on the second term for the inequality. We compute \( \phi^{(\ell)}(t) \) as
\[
\phi^{(\ell)} = \left( \phi^{(\ell-1)} \right)^2 \phi \circ \phi^{(\ell-1)} + \phi^{(\ell-1)} \phi \circ \phi^{(\ell-1)}
\]
and thus
\[
\frac{\phi^{(\ell)}}{\phi^{(\ell)}} = \phi^{(\ell-1)} \phi \circ \phi^{(\ell-1)} + \phi^{(\ell-1)} \phi^{(\ell-1)} = \sum_{\ell' = 0}^{\ell-1} \phi^{(\ell')} \phi \circ \phi^{(\ell')},
\]
which gives
\[
|\phi^{(\ell)}| = \left| \phi^{(\ell)} \sum_{\ell' = 0}^{\ell-1} \phi^{(\ell')} \phi \circ \phi^{(\ell')} \right|.
\]
From Lemma F.14, we have \( |\phi| \leq c_1 = 4 \) on \( t \in [0, \pi] \) and \( \phi \geq c_2 = \frac{1}{3} \) on \( [0, \frac{t}{2}] \). As when \( \ell > 0 \), we have \( \phi^{(\ell)}(t) \leq \frac{t}{2} \), we separate the case when \( \ell = 0 \). From Lemma F.16 we get
\[
\sum_{\ell = 0}^{\infty} \phi^{(\ell)}(t) \leq \frac{C}{t}.
\]
Using the chain rule to get the expression for \( \phi^{(\ell)} \), and concavity of \( \phi \) to get that \( \phi(t) \geq t/2 \), and decreasingness of \( \phi \), we have
\[
|\phi^{(\ell)}(t)| \leq |\phi(t)| \prod_{\ell' = 1}^{\ell-1} \phi \circ \phi^{(\ell')}(t) + \frac{c_1}{c_2} \phi^{(\ell)}(t) \sum_{\ell' = 1}^{\ell-1} \phi^{(\ell')}(t)
\]
\[
\leq c_1 \prod_{\ell' = 1}^{\ell-1} \phi \circ \phi^{(\ell')}(t) + \frac{c_1}{c_2} \phi^{(\ell)}(t) \sum_{\ell' = 1}^{\ell-1} \phi^{(\ell')}(t)
\]
\[
\leq c_1 \prod_{\ell' = 0}^{\ell-2} \phi \circ \phi^{(\ell')} \circ \phi(t) + \frac{C}{t} \phi^{(\ell)}(t)
\]
\[
\leq c_1 \prod_{\ell' = 0}^{\ell-2} \phi \circ \phi^{(\ell')}(t/2) + \frac{C}{t} \phi^{(\ell)}(t)
\]
\[
\leq 2c_1 \prod_{\ell' = 0}^{\ell-1} \phi \circ \phi^{(\ell')}(t/2) + \frac{C}{t} \phi^{(\ell)}(t/2)
\]
\[
\leq 8 \phi^{(\ell)}(t/2) + \frac{C}{t} \phi^{(\ell)}(t/2)
\]
\[
\leq \frac{C}{t} \phi^{(\ell)}(t/2). \tag{17}
\]
From Lemma G.5, we know \( \xi_1 \) is monotonically decreasing, so \( \xi_\ell(t) \leq \xi_\ell(t/2) \). Thus (proceeding from our previous bound)
\[
\frac{\xi_\ell(t)}{\pi} \sum_{\ell' = \max(1, \ell)}^{L-1} \phi^{(\ell')}(t) 1 - \phi^{(\ell')}(t)/\pi \leq \frac{C}{t} \left| \frac{\xi\ell(t/2)}{\pi} \sum_{\ell' = \max(1, \ell)}^{L-1} \phi^{(\ell')}(t/2) 1 - \phi^{(\ell')}(t/2)/\pi \right|
\]
\[
\leq \frac{2C}{t} \left| \frac{\xi\ell(t/2)}{\pi} \sum_{\ell' = \max(1, \ell)}^{L-1} \phi^{(\ell')}(t/2) 1 - \phi^{(\ell')}(t/2)/\pi \right|
\]
\[
\leq \frac{2C}{t} \left| \frac{\xi\ell(t/2)}{\pi} \sum_{\ell' = \max(1, \ell)}^{L-1} \phi^{(\ell')}(t/2) 1 - \phi^{(\ell')}(t/2)/\pi \right|.
\]
75
where in the second line we used that $1/2 \leq 1 - \phi(t)/\pi \leq 1$ and the $t/2$ term is no smaller. Similarly, applying Lemma F.16 again, we get

$$\left| \frac{2\xi(t)}{\pi^2} \mathbb{1}_{\ell=0} \sum_{\ell'=1}^{L-1} \frac{\dot{\phi}^{(|\ell'|)}(t)}{1 - \phi^{(|\ell'|)}/\pi} \right| \leq C \frac{\xi_1(t)}{t} \mathbb{1}_{\ell=0} \leq C \frac{\xi_1(t/2)}{t} \mathbb{1}_{\ell=0}$$

and

$$\frac{\xi(t)}{\pi^2} \left( \sum_{\ell'=\max\{1,\ell\}}^{L-1} \frac{\dot{\phi}^{(|\ell'|)}(t)}{1 - \phi^{(|\ell'|)}/\pi} \right)^2 \leq \xi(t) \left( \frac{1}{\ell t} \frac{1}{1 + \ell t/(3\pi)} \right)^2 \leq C \frac{\xi(t)}{\ell^2} \left( 1 + \ell t/(3\pi) \right)^{-2}$$

Combining all these bounds and applying Lemma G.9, we have obtained

$$\left| \dot{\xi}(t) \right| \leq \frac{C}{\ell^2} \left| \dot{\xi}(t/2) \right| + \frac{C'}{\ell^2} \frac{1}{1 + \ell t/\pi}$$

$$\leq \frac{1}{\ell^2} \left| \dot{\xi}(t/2) \right| + \frac{1}{\ell^3}.$$

Note this holds for all $\ell = 0, \ldots, L - 1$, so we directly get $|\ddot{\psi}(t)| \leq \frac{C}{\ell} |\dot{\psi}(t/2)| + \frac{nL}{\ell^2} C'$. Thus from Lemma F.10, there exists constant $C, C'_1, C''_1$, when $L \geq C$, we have

$$\max_{\ell \geq r} |\ddot{\psi}(t)| \leq \max_{\ell \geq r} \left( \frac{C'}{\ell} |\dot{\psi}(t/2)| + \frac{nL C'}{\ell^3} \right)$$

$$\leq \frac{1}{r} \left( \frac{C'}{r^2} \right)^n + C' \frac{n}{r^3},$$

which provides the bound for $L^\infty$ control. For $L^1$ control, we have

$$\int_{t=r}^{\pi} t^6 |\ddot{\psi}(t)| dt \leq \int_{t=0}^{\pi} t^6 \left( \frac{C}{t} |\dot{\psi}(t)| + \frac{nL C'}{\ell t^3} \right) dt$$

$$\leq C \int_{t=r}^{\pi} t^3 |\dot{\psi}(t/2)| dt + C'n$$

$$\leq Cn,$$

which finishes the proof.

**Lemma F.12.** There exist absolute constants $C, C'$ such that when $L \geq C$ we have for any $r > 0$

$$\max_{\ell \geq r} |\ddot{\psi}(t)| \leq \frac{C'n}{r^4} \tag{F.18}$$

and

$$\int_{t=0}^{\pi} t^6 |\ddot{\psi}(t)| dt \leq C'n.$$

**Proof.** We calculate with the chain rule starting from the representation in Lemma G.8 (and use the triangle inequality)

$$\left| \dot{\xi}_t \right| \leq \frac{\xi(t)}{\pi} \sum_{\ell'=\max\{1,\ell\}}^{L-1} \frac{\dot{\phi}^{(|\ell'|)}(t)}{1 - \phi^{(|\ell'|)}/\pi} + \frac{2\xi(t)}{\pi^2} \mathbb{1}_{\ell=0} \sum_{\ell'=1}^{L-1} \frac{\dot{\phi}^{(|\ell'|)}(t)}{1 - \phi^{(|\ell'|)}/\pi}$$

$$+ \frac{\xi(t)}{\pi^3} \sum_{\ell', \ell'' = \max\{1,\ell\}}^{L-1} \left( \frac{\dot{\phi}^{(|\ell'|)\phi^{(|\ell'|)}}(t)}{1 - \phi^{(|\ell'|)}/\pi} \right)$$

$$+ \frac{\xi(t)}{\pi^3} \sum_{\ell', \ell'' = \max\{1,\ell\}}^{L-1} \left( \frac{\dot{\phi}^{(|\ell'|)\phi^{(|\ell'|)}}(t)}{1 - \phi^{(|\ell'|)}/\pi} \right)$$
where the second line uses induction. From Lemmas F.14 and F.15, we have

\[
\left| \frac{\partial}{\partial \ell} \right| \leq 2C/t \sum_{\ell' = 1}^{L-1} \frac{|\varphi[\ell'](t)|}{1 - \varphi[\ell'](t)/\pi}
\]

which leaves the main unresolved term in (F.19) to be \( \bar{\varphi} \). On the other hand, we have from the chain rule

\[
\bar{\varphi}[\ell] = 3\bar{\varphi}[\ell-1] \varphi[\ell-1](\bar{\varphi} \circ \varphi[\ell-1]) + \left( \varphi[\ell-1] \right)^3 \left( \bar{\varphi} \circ \varphi[\ell-1] \right) + \bar{\varphi}[\ell-1] \left( \bar{\varphi} \circ \varphi[\ell-1] \right).
\]

Using the product expression \( \bar{\varphi}[\ell] = \varphi[\ell-1] \bar{\varphi} \circ \varphi[\ell-1] \) and the triangle inequality, we have

\[
\left| \frac{\bar{\varphi}[\ell]}{\bar{\varphi}[\ell-1]} \right| \leq \left| \frac{\varphi[\ell-1]}{\bar{\varphi}[\ell-1]} \right| \left| \bar{\varphi} \circ \varphi[\ell-1] \right| + \left( \varphi[\ell-1] \right)^2 \left| \bar{\varphi} \circ \varphi[\ell-1] \right| + \left( \bar{\varphi}[\ell-1] \right) \left| \varphi[\ell-1] \right|
\]

where the second line uses induction. From Lemmas F.14 and F.15, we have \( |\bar{\varphi}| \leq c_1 = 4, |\bar{\varphi}| \leq c_4 \) on \( t \in [0, \pi] \) and \( \varphi \geq c_2 = 1/2, \bar{\varphi} \leq -c_3 \) on \( [0, \pi/2] \). Again, for \( \ell > 0 \), we have \( \varphi[\ell](t) \leq \pi/2 \). Applying (F.20) and Lemma F.16, we get

\[
\frac{|\varphi[\ell](t)|}{\varphi[\ell](t)} \leq 3 \frac{c_1}{c_2} \sum_{\ell' = 1}^{L} |\varphi[\ell']| + \frac{c_4}{c_2} \sum_{\ell' = 1}^{L} \left( \varphi[\ell'](t) \right)^2 + \frac{c_4}{\varphi(t)}
\]

Multiplying both side with \( \varphi[\ell] \), we get the bound

\[
\left| \varphi[\ell](t) \right| \leq \frac{C}{t^2} \varphi[\ell](t) + \frac{c_4}{c_2} \prod_{\ell' = 1}^{\ell-1} \bar{\varphi} \circ \varphi[\ell'](t)
\]

\[
\leq \frac{C}{t^2} \varphi[\ell](t) + \frac{c_4}{c_2} \prod_{\ell' = 0}^{\ell-2} \bar{\varphi} \circ \varphi[\ell'] \circ \varphi(t)
\]

\[
\leq \frac{C}{t^2} \varphi[\ell](t) + 2c_4 \varphi[\ell](t/2)
\]

\[
\leq \frac{C}{t^2} \varphi[\ell](t/2),
\]

77
where the justifications for this argument are very similar to those used in the proof of Lemma F.14.

Plugging bounds we have here back to (F.19). From Lemma F.16 and monotonicity of $\xi_\ell$ in Lemma G.5, we get

\[
\left| \partial_\ell \xi(t) \right| \leq \left| \frac{\xi(t)}{\pi} \sum_{\ell' = \max\{1, \ell\}}^{L-1} \frac{C_\ell \varphi[\ell](t/2)}{1 - \varphi[\ell](t/\pi)} \right| \\
+ \frac{\xi(t) C}{t^2} \sum_{\ell' = \max\{1, \ell\}}^{L-1} \frac{\varphi[\ell](t)}{1 - \varphi[\ell](t/\pi)} \left( \frac{C_\ell}{\ell^2} \right)^2 \sum_{\ell' = \max\{1, \ell\}}^{L-1} \frac{\varphi[\ell](t)}{1 - \varphi[\ell](t/\pi)} \\
+ \frac{\xi(t) C}{t^2} \sum_{\ell' = \max\{1, \ell\}}^{L-1} \frac{\varphi[\ell](t/2)}{1 - \varphi[\ell](t/2)/\pi} \left( \frac{C_\ell}{\ell^2} \right)^2 \sum_{\ell' = \max\{1, \ell\}}^{L-1} \frac{\varphi[\ell](t)}{1 - \varphi[\ell](t/\pi)}
\]

where from the second to third line we also use the fact that $1/2 \leq 1 - \varphi[\ell](t)/\pi \leq 1$ for all $\ell \geq 1$ and the last line follows from the formula of $\xi_\ell$ in (F.16).

From our bounds of $\psi(t)$ in (F.13), this leads to

\[
\max_{t \geq r} \left| \tilde{\psi}(t) \right| \leq \max_{t \geq r} \left| \frac{C}{t^2} \psi(t/2) \right| \\
\leq C \frac{C'}{r^2 (r/2)^2} \\
= Cn \frac{1}{r^4}
\]

and

\[
\int_{t=0}^{\pi} t^9 \tilde{\psi}(t) dt \leq \int_{t=0}^{\pi} t^9 \frac{C}{t^2} \psi(t/2) dt \\
\leq Cn
\]

as claimed.

\[ \square \]

F.5 Additional Proofs for Some Bounds

**Lemma F.13.** There exists an absolute constant $C_1 > 0$ such that

\[
\tilde{\varphi}(t) - \varphi(t) \leq C_1 t^3.
\]

**Proof.** From Lemma G.4, $\varphi$ is 3 times continuously differentiable on $(0, \pi)$, and

\[
\varphi(0) = 0, \quad \varphi(0) = 1, \quad \varphi(0) = -\frac{2}{3\pi}.
\]

It is easy to check that

\[
\tilde{\varphi}(0) = 0, \quad \tilde{\varphi}(0) = 1, \quad \tilde{\varphi}(0) = -\frac{2}{3\pi}.
\]

\[ 78 \]
Since the Taylor expansions of these two functions around 0 agree to third order, and both are 3 times continuously differentiable on \((0, \pi)\), we obtain by Lagrange’s remainder theorem that for any \(t \in [0, \pi)\),

\[
\hat{\varphi}(t) - \varphi(t) = \int_0^t \left( \hat{\varphi}(s) - \varphi(s) \right) \frac{s^2}{2} \, ds \leq C_1 t^3
\]

for some finite constant \(C_1 = \sup_{t \in [0, \pi)} |\hat{\varphi}(t) - \varphi(t)|\). At \(t = \pi\) we have \(\hat{\varphi}(\pi) - \varphi(\pi) = \frac{\pi}{\pi + \pi/3} - \frac{\pi}{3} \leq 0\) hence the same bound holds for \(t \in [0, \pi]\).

**Lemma F.14.** One has

\[
\varphi(t) \geq \frac{1}{2}, \quad t \in [0, \frac{\pi}{2}]
\]

\[
|\varphi(t)| \leq 4, \quad t \in [0, \pi]
\]

**Proof.** We know \(\varphi\) is monotonically increasing and concave on \([0, \pi]\), thus for \(t \in [0, \pi/2]\),

\[
\varphi(t) \geq \frac{\varphi(\pi/2)}{1/2} \geq \frac{1}{2}.
\]

Using Lemma G.6 we also have for \(t \in [0, \pi]\), \(\dot{\varphi}(t) \leq \dot{\varphi}(0) = 1\),

\[
\varphi(t) \geq \frac{t}{1 + t/\pi} \geq \frac{t}{2},
\]

and the first bound here can be used to obtain

\[
t - \varphi(t) \leq \frac{t^2/\pi}{1 + t/\pi} \leq t^2/\pi.
\]

Thus since \(\varphi \leq \pi/2\)

\[
\cos t \sin \varphi(t) - \varphi(t) \sin \cos \varphi(t) \geq \cos t \sin \varphi(t) - \sin t \cos \varphi(t) \geq -\sin(t - \varphi(t)),
\]

and in particular, using the expression for \(\ddot{\varphi}\) from Lemma G.4

\[
-\ddot{\varphi}(t) = -(1 - \frac{t}{\pi}) \cos t \sin \varphi(t) - \varphi(t) \sin t \cos \varphi(t) \leq \frac{\sin t}{\pi \sin(\varphi(t))} + \frac{\sin t}{\pi \sin(\varphi(t))} \leq \frac{t^2/\pi}{\sin^2(t/2)} + \frac{2}{\pi} \leq \frac{t^2/\pi}{(t/\pi)^2} + \frac{2}{\pi} \leq 4.
\]

**Lemma F.15.** There exist constants \(c_3, c_4 > 0\) such that \(\ddot{\varphi}(t) < -c_3\) for \(t \in [0, \pi/2]\) and \(|\dddot{\varphi}| \leq c_4\) for \(t \in [0, \pi]\).

**Proof.** The existence of \(c_3\) follows from Lemma G.4 directly. The existence of \(c_4\) follows from smoothness of \(\varphi\) on \((0, \pi)\) and the fact that \(\dddot{\varphi}(0) = -\frac{1}{\pi^2}, \dddot{\varphi}(\pi) = \frac{2}{\pi}\) both exist.
Lemma F.16. There exists an absolute constant $C > 0$ such that for any $0 < t \leq \pi$ and $\ell \in \mathbb{N}_0$, one has

$$\sum_{\ell' = \ell}^{\infty} \varphi[\ell'](t) \leq \frac{C}{\ell} \frac{1}{1 + \ell t/(3\pi)}.$$

Proof. Using Lemma F.3, we have

$$\varphi[\ell](t) \leq \frac{C}{(1 + \ell t/(3\pi))^2}.$$

We can then calculate

$$\sum_{\ell' = \ell}^{\infty} \varphi[\ell'](t) \leq C \sum_{\ell' = \ell}^{\infty} \frac{1}{(1 + \ell t/(3\pi))^2} \leq C \left( \frac{1}{1 + \ell t/(3\pi)} + \int_{\ell' = \ell}^{\infty} \frac{1}{(1 + \ell t/(3\pi))^2} \, d\ell' \right) \leq C \left( \frac{1}{1 + \ell t/(3\pi)} + \frac{3\pi/t}{1 + \ell t/(3\pi)} \right) \leq C \frac{1}{\ell} \frac{1}{1 + \ell t/(3\pi)},$$

as claimed. 

G Auxiliary Results

Results in this section are reproduced from the literature for self-containedness, and for the most part are presented without proofs.

G.1 Certificates Imply Generalization

Theorem G.1 ([6, Theorem B.1], specialized slightly). Let $\mathcal{M}$ be a two curve problem instance. For any $0 < \delta \leq 1/e$, choose $L$ so that

$$L \geq C_1 \max\{C_{\mu} \log^{9}(1/\delta) \log^{24}(C_{\mu} \log(1/\delta)), \kappa^2 C_{\lambda}\},$$

let $N \geq L^{10}$, set $n = C_2 L^{90} \log^{9}(1/\delta) \log^{18}(Ln_0)$, and fix $\tau > 0$ such that

$$\frac{C_3}{nL^{2}} \leq \tau \leq \frac{C_4}{nL^2}.$$

Then if there exists a function $g \in L^2_{\mu}\infty(\mathcal{M})$ such that

$$\|\Theta^\mu_{\text{NTK}}[g] - \zeta_0\|_{L^2_{\mu}\infty(\mathcal{M})} \leq C_5 \sqrt{\log(1/\delta) \log(nn_0)} \frac{L \min\{\rho_{\text{cert}}^{-q}, \rho_{\text{min}}^{-q}\}}{\rho_{\text{cert}}^{2}}, \quad \|g\|_{L^2_{\mu}\infty(\mathcal{M})} \leq C_6 \sqrt{\log(1/\delta) \log(nn_0)} \frac{1}{n_{\text{cert}}^{2}},$$

with probability at least $1 - \delta$ over the random initialization of the network and the i.i.d. sample from $\mu$, the parameters obtained at iteration $|L^{30}/44/(n\tau)|$ of gradient descent on the finite sample loss $L_{\mu\nu}$ yield a classifier that separates the two manifolds.

The constants $C_1, \ldots, C_4 > 0$ depend only on the constants $q_{\text{cert}}, C_5, C_6 > 0$, the constants $\kappa, C_{\lambda}$ are respectively the extrinsic curvature constant and the global regularity constant defined in [6, §2.1], and the constant $C_{\mu}$ is defined as $\max\{\rho_{\text{min}}^{-q}, \rho_{\text{min}}^{-q}\}(1 + \rho_{\text{max}})^6 (\min\{\mu(\mathcal{M}_+), \mu(\mathcal{M}_-))\}^{-11/2}$, where $q = 11 + 8q_{\text{cert}}$.

G.2 Concentration of the Initial Random Network and Its Gradients

Theorem G.2 (Corollary of [6, Theorem B.2, Lemma C.11]). Let $\mathcal{M}$ be a two curve problem instance. For any $d \geq K \log(nn_0 \log(\log n))$, if $n \geq K'd^4 L^5$ then one has on an event of probability at least $1 - e^{-cd}$

$$\|\Theta - \Theta^\mu_{\text{NTK}}\|_{L^\infty(\mathcal{M} \times \mathcal{M})} \leq Cn/L,$$

where $c, C, K, K' > 0$ are absolute constants.
Lemma G.3 ([6, Lemma D.11]). There are absolute constants $K, K' > 0$ such that if $d \geq K \log(n_0 \operatorname{len}(M))$ and $n \geq K'd^4L$, then
\[
\begin{align*}
\mathbb{P} \left[ \|f_{\theta_0}\|_{L^\infty} \leq \sqrt{d} \right] &\geq 1 - e^{-cd}, \\
\mathbb{P} \left[ \|\zeta_0\|_{L^\infty} \leq \sqrt{d} \right] &\geq 1 - e^{-cd}.
\end{align*}
\]
Define
\[
\zeta(x) = -f_*(x) + \int_{M} f_{\theta_0}(x')d\mu(x').
\]
Then under the same assumptions
\[
\begin{align*}
\mathbb{P} \left[ \|\zeta_0 - \zeta\|_{L^\infty} \leq \sqrt{\frac{d}{L^2}} + \sqrt{\frac{L}{n}} \right] \geq 1 - e^{-cd}
\end{align*}
\]
for some numerical constant $c$.

G.3 Basic Estimates for the Infinite-Width Neural Tangent Kernel

Lemma G.4 ([6, Lemma E.5]). One has

1. $\phi \in C^\infty(0, \pi)$, and $\phi^\prime$, $\phi^\prime\prime$, and $\phi^\prime\prime\prime$ extend to continuous functions on $[0, \pi]$;
2. $\phi(0) = 0$ and $\phi(\pi) = \pi/2$; $\phi^\prime(0) = 1$, $\phi^\prime(0) = -2/(3\pi)$, and $\phi^\prime\prime(\pi) = -1/(3\pi^2)$; and $\phi^\prime(\pi) = \phi^\prime(\pi) = 0$;
3. $\phi$ is concave and strictly increasing on $[0, \pi]$ (strictly concave in the interior);
4. $\phi^\prime < -c < 0$ for an absolute constant $c > 0$ on $[0, \pi/2]$;
5. $0 < \phi^\prime < 1$ and $0 > \phi^\prime \geq -C$ on $[0, \pi)$ for some absolute constant $C > 0$;
6. $\nu(1 - C_1\nu) \leq \nu(\nu) \leq \nu(1 - c_1\nu)$ on $[0, \pi]$ for some absolute constants $C_1, c_1 > 0$.

**Proof.** Combine the results in [6, Lemma E.5] with Lemma F.15 to obtain the conclusion. □

Lemma G.5 (Corollaries of Lemma G.4, stated in [6, Lemma C.10]). One has:

1. The function $\phi$ is smooth on $(0, \pi)$, and (at least) $C^3$ on $[0, \pi]$.
2. For each $\ell = 0, 1, \cdots, L$, the functions $\phi^{[\ell]}$ are nonnegative, strictly increasing, and concave (positive and strictly concave on $[0, \pi)$).
3. If $0 \leq \ell < L$, the functions $\xi_{\ell}$ are nonnegative, strictly decreasing, and convex (positive and strictly convex on $[0, \pi)$).
4. The function $\phi$ is smooth on $(0, \pi)$, $C^3$ on $[0, \pi]$, and is nonnegative, strictly decreasing, and convex.

Lemma G.6 ([6, Lemma C.13]). If $\ell \in \mathbb{N}_0$, the iterated angle evolution function satisfies the estimate
\[
\phi^{[\ell]}(t) \geq \frac{t}{1 + \ell t/\pi},
\]

Lemma G.7 ([6, Lemma C.17]). One has for every $\ell \in \{0, 1, \cdots, L\}$
\[
\begin{align*}
\phi^{[\ell]}(0) &\geq 0; \quad \phi^{[\ell]}(0) = 1; \quad \phi^{[\ell]}(0) = -2\ell/3\pi,
\end{align*}
\]
and for $\ell \in [L]$,
\[
\phi^{[\ell]}(\pi) = \phi^{[\ell]}(\pi) = 0.
\]
Finally, we have $\phi^{[0]}(\pi) = 1$ and $\phi^{[0]}(\pi) = 0$. 

81
Lemma G.8 ([6, Lemma C.18]). For first and second derivatives of $\xi_\ell$, one has

$$\dot{\xi}_\ell = -\pi^{-1} \sum_{\ell' = \ell}^{L-1} \phi[\ell'] \prod_{\ell'' = \ell}^{L-1} (1 - \pi^{-1} \phi[\ell'']),$$

and

$$\ddot{\xi}_\ell = -\frac{1}{\pi} \sum_{\ell' = \ell}^{L-1} \left[ \phi[\ell'] \prod_{\ell'' = \ell}^{L-1} (1 - \pi^{-1} \phi[\ell'']) - \pi^{-1} \phi[\ell'] \sum_{\ell''' = \ell}^{L-1} \phi[\ell''] \prod_{\ell'' = \ell}^{L-1} (1 - \pi^{-1} \phi[\ell''']) \right],$$

where empty sums are interpreted as zero, and empty products as 1. In particular, one calculates

$$\xi_\ell(0) = 1; \quad \dot{\xi}_\ell(0) = -\frac{L - \ell}{\pi}; \quad \ddot{\xi}_\ell(0) = \frac{(L - \ell)(L - \ell - 1)}{\pi^2} + \frac{L(L - 1) - \ell(\ell - 1)}{3\pi^2},$$

and

$$\xi_0(\pi) = 0; \quad \dot{\xi}_\ell(\pi) = -\frac{1}{\pi} \xi_1(\pi) \mathbb{1}_{\ell=0}; \quad \ddot{\xi}_\ell(\pi) = 0.$$

Lemma G.9 ([6, Lemma C.20]). For all $\ell \in \{0, 1, \ldots, L - 1\}$, one has

$$\xi_\ell(t) \leq \frac{1 + \ell t / \pi}{1 + Lt / \pi}.$$

Lemma G.10 ([6, Lemma C.21]). One has

$$|\dot{\xi}_\ell(t)| \leq \frac{3 (L - \ell)}{1 + Lt / \pi}.$$

Lemma G.11 ([6, Lemma C.23]). There are absolute constants $c, C > 0$ such that for all $\ell \in \{0, \ldots, L - 1\}$, one has

$$\left| \ddot{\xi}_\ell \right| \leq C \frac{L(L - \ell)(1 + \ell \nu / \pi)}{(1 + cL\nu)^2} + C \frac{(L - \ell)^2}{(1 + cL\nu)(1 + c\ell\nu)}.$$

References