Appendices

A. Implementation details

A.1. Tangent-space optimization

Due to manifold constraints, rotations cannot be naively optimized using standard first-order optimizers. In TILTED, we address this via a Riemannian ADAM [17] approach. Each $\boldsymbol{\tau}_t$ is stored as a unit-complex vector ($\in \mathbb{S}^1$) for 2D experiments and as a unit quaternion ($\in \mathbb{S}^3$) for 3D experiments, but gradients are computed with respect to tangent spaces corresponding to the standard $\mathfrak{so}(2)$ and $\mathfrak{so}(3)$ Lie algebras. ADAM [13] is applied to scale tangent-space gradients ξ_t^k at each training step k, and an exponential map is used in place of addition to apply updates:

$$\boldsymbol{\tau}_{t,k+1} = \boldsymbol{\tau}_{t,k}\text{Exp}(\alpha_k \xi_{t,k})$$

where α_k is the learning rate for $\boldsymbol{\tau}$ at step k. For experiments with real world data, we refine camera poses using this same mechanism.

A.2. Handling boundaries

One benefit of axis-aligned latent decompositions is that they make bounding boxes intuitive: all coordinates used for interpolation can be constrained to lie within a well-defined input domain. When we apply geometric transformations to the domain of factors, however, the regions of the input space that each factor covers stop fully overlapping. To resolve this for bounded scenes, we apply simple coordinate clipping. Toroidal boundary conditions, similar to what is used in Factor Fields [34], can also be used. For unbounded scenes, we adopt an ℓ^∞ norm-based scene contraction function [24, 36]:

$$\text{contract}(\boldsymbol{p}) = \begin{cases} \boldsymbol{p} & ||\boldsymbol{p}||_\infty \leq 1 \\ (2 - \frac{1}{||\boldsymbol{p}||_\infty})(\frac{\boldsymbol{p}}{||\boldsymbol{p}||_\infty}) & ||\boldsymbol{p}||_\infty > 1 \end{cases} \tag{A.1}$$

When applied *after* $\boldsymbol{\tau}$, note that scene contraction places all points in the range $[-2, 2]$, which mitigates boundary concerns entirely.

A.3. Regularization

We adopt two standard regularization terms: spatial total variation (TV) on feature grids and the distortion loss proposed by MipNeRF 360 [24]. NeRF experiments additionaly rely on a pair of proposal fields, which require an additional interlevel loss [24]. A weight of 0.01 is used for total variation, a weight of 0.001 for the distortion loss, and a weight of 1.0 for the interlevel loss. More details can be found in our code release.

B. Disaggregated SDF results

B.1. SDF results, with random scene rotation

In this section, we report disaggregated results from our SDF reconstruction experiments, with and without TILTED. We apply a random global rotation for each seed in these results.

Methods	Avg	Bunny$^{SO(3)}$	Lucy$^{SO(3)}$	Chair$^{SO(3)}$	Armadillo$^{SO(3)}$	Dragon$^{SO(3)}$	Cheburashka$^{SO(3)}$	Beast$^{SO(3)}$	Happy$^{SO(3)}$
IoU↑									
K-Planes-30	0.949	0.969	0.933	0.937	0.952	0.935	0.980	0.922	0.967
w/ TILTED	0.989	0.996	**0.987**	0.987	0.993	0.977	0.995	0.988	0.990
K-Planes-60	0.949	0.982	0.939	0.922	0.955	0.922	0.979	0.918	0.978
w/ TILTED	0.990	0.996	0.982	**0.993**	0.989	0.983	**0.997**	0.984	**0.993**
K-Planes-90	0.946	0.967	0.951	0.898	0.946	0.929	0.989	0.913	0.974
w/ TILTED	**0.991**	**0.996**	0.981	0.991	**0.994**	**0.986**	0.995	**0.990**	0.992

Table 1: **K-Plane results for SDF reconstruction *with* random scene rotation.** We report metrics with 30, 60, and 90 channels.

Methods	Avg	Bunny	Lucy	Chair	Armadillo	Dragon	Cheburashka	Beast	Happy
IoU↑									
VM-45	0.866	0.974	0.802	0.950	0.913	0.821	0.969	0.977	0.519
w/ TILTED	0.974	0.994	0.973	0.936	0.988	0.952	0.979	0.981	0.991
VM-90	0.946	0.982	0.956	0.948	0.984	0.762	0.981	0.972	0.985
w/ TILTED	0.977	0.995	**0.984**	0.897	**0.994**	0.978	**0.995**	0.987	0.989
VM-135	0.982	0.988	0.969	0.974	0.987	0.971	0.986	**0.988**	0.991
w/ TILTED	**0.988**	**0.996**	0.982	**0.976**	0.992	**0.981**	0.994	0.987	**0.994**

Table 2: **Vector-matrix results for SDF reconstruction *with* random scene rotation.** We report metrics with 45, 90, and 135 channels.

B.2. SDF results, without random scene rotation

In this section, we report SDF reconstruction metrics when we turn off random scene rotation. Metrics here are similar to those when we include random scene rotation. In the main paper body, we report metrics with random rotation included.

Methods	Avg	Bunny	Lucy	Chair	Armadillo	Dragon	Cheburashka	Beast	Happy
IoU↑									
K-Planes-30	0.949	0.970	0.945	0.965	0.945	0.843	0.989	0.970	0.966
w/ TILTED	0.989	0.996	0.983	0.988	0.992	0.979	0.995	0.988	0.990
K-Planes-60	0.952	0.972	0.954	0.964	0.951	0.842	0.993	0.972	0.969
w/ TILTED	0.990	**0.997**	0.982	**0.991**	0.991	**0.981**	**0.996**	0.989	**0.993**
K-Planes-90	0.952	0.977	0.945	0.959	0.961	0.838	0.994	0.971	0.971
w/ TILTED	**0.991**	0.996	**0.985**	0.990	**0.996**	0.979	0.994	**0.995**	0.992

Table 3: **K-Plane results for SDF reconstruction *without* random scene rotation.** We report metrics with 30, 60, and 90 channels.

Methods	Avg	$\text{Bunny}^{SO(3)}$	$\text{Lucy}^{SO(3)}$	$\text{Chair}^{SO(3)}$	$\text{Armadillo}^{SO(3)}$	$\text{Dragon}^{SO(3)}$	$\text{Cheburashka}^{SO(3)}$	$\text{Beast}^{SO(3)}$	$\text{Happy}^{SO(3)}$
IoU↑									
VM-45	0.970	0.990	0.927	0.975	0.970	0.952	0.988	0.981	0.980
w/ TILTED	0.982	0.995	0.980	0.980	0.970	0.975	0.988	0.982	0.989
VM-90	0.979	0.993	0.971	**0.991**	0.955	0.960	0.992	0.983	0.988
w/ TILTED	**0.989**	0.995	0.985	0.989	0.993	0.976	0.993	**0.987**	0.991
VM-135	0.982	0.993	0.973	0.987	0.991	0.964	0.977	0.981	0.989
w/ TILTED	0.988	**0.996**	**0.985**	0.989	**0.994**	**0.983**	**0.997**	0.966	**0.993**

Table 4: **Vector-matrix results for SDF reconstruction *without* random scene rotation.** We report metrics with 45, 90, and 135 channels.

B.3. Ablations on coarse-to-fine optimization

We report an ablation for the low pass-based coarse-to-fine optimization in Table 5.

Methods (KPlane)	Axis-aligned	with TILTED-5	with TILTED-10
Runtime (minutes : seconds)	06:29	08:03	14:05

Table 6: SDF KPlane Runtime

Methods	Lucy	Dragon
$IoU \uparrow$		
TILTED-VM w/o coarse-to-fine	0.975	0.976
TILTED-VM w/ coarse-to-fine	**0.985**	**0.981**
TILTED-K-Planes w/o coarse-to-fine	0.974	0.977
TILTED-K-Planes w/ coarse-to-fine	**0.988**	**0.979**

Table 5: **Ablation for coarse-to-fine optimization inspired by Nerfies [27] and BARF [26].** Coarse-to-fine optimization consistently improves performance for TILTED SDF reconstructions.

C. 2D Results

C.1. Experiments on various latent grid resolutions

In this section, we vary latent grid resolution for 2D image regression task. As shown in Figure 1, with large grid resolution, axis-aligned models will produce noisy images whereas TILTED can produce images with much less noise.

Grid Resolution	32	64	128	256	512	1024
Fox (axis-aligned)	21.26	21.98	22.31	21.63	17.23	10.34
Fox (TILTED)	**21.33**	**22.19**	**22.52**	**22.23**	**19.21**	**17.00**
Bridge (axis-aligned)	20.95	21.96	22.99	23.63	23.46	20.90
Bridge (TILTED)	**21.43**	**22.28**	**23.34**	**24.16**	**24.08**	**22.23**
Painting (axis-aligned)	25.59	26.16	26.51	26.76	26.40	18.22
Painting (TILTED)	**25.83**	**26.5**	**26.81**	**26.94**	**26.81**	**22.15**

Table 7: PSNR for 2D Image Regression task using various latent grid resolutions

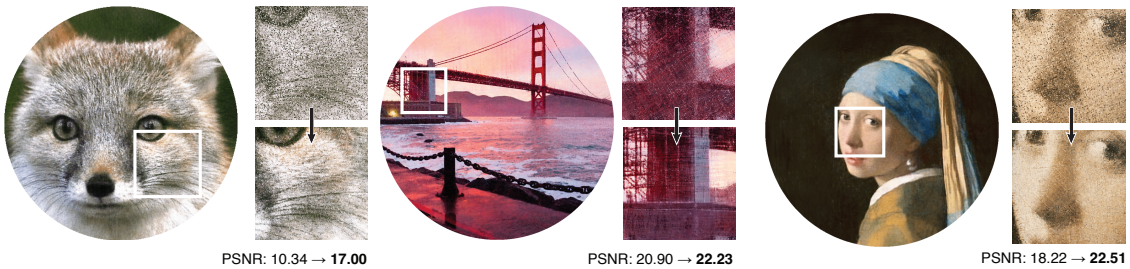

Figure 1: **Implicit regularization via TILTED.** Reconstruction quality improves dramatically when TILTED is applied to an overparameterized latent grid with resolution 1024.

D. 2D Runtime Analysis

We compare and report runtime for 2D image regression. We train axis-aligned and TILTED model variants with eight transformations for 5K iterations. Experiments are run on a single RTX3080 GPU.

Methods	Axis-aligned	with TILTED-8
Runtime (minutes : seconds)	03:24	03:26

Table 8: **2D runtime measurement.**

E. Unifying factored feature volumes

In this section, we concretize how feature volume decompositions used by prior work can be instantiated using the common notation that we present:

$$\boldsymbol{Z} = \texttt{Reduce}\big(\big[\texttt{Interp}_{\boldsymbol{F}_1}(\texttt{Proj}_1(\boldsymbol{p}))\big], \ldots, \big[\texttt{Interp}_{\boldsymbol{F}_F}(\texttt{Proj}_F(\boldsymbol{p}))\big]\big), \tag{E.1}$$

where, as earlier, $\boldsymbol{p}$ is an input coordinate and $\boldsymbol{Z}$ is an output that can be used to regress quantities like radiance or signed distance. This unified formulation, which closely mirrors the structure of our implementation, enables integration of the latent registration mechanism proposed by TILTED in a general-purpose way.

E.1. Vector outer products

Among the best-known approaches for factoring tensors is the classic CANDECOMP/PARAFAC (CP) decomposition, which has been studied as a baseline for factoring latent grids in prior work [31]. In 3D, the CP decomposition is equivalent to a single vector-matrix decomposition when the matrix rank is constrained to rank-1 and can thus be represented with a vector outer product.

To build CP-decomposed latent structures, a channel dimension is included to instantiate three paired 1D feature grids and projection functions:

$$\begin{aligned}
\boldsymbol{F}_1 &\in \mathbb{R}^{w \times c} \quad \texttt{Proj}_1(\boldsymbol{p}) = p_x \in \mathbb{R} \\
\boldsymbol{F}_2 &\in \mathbb{R}^{h \times c} \quad \texttt{Proj}_2(\boldsymbol{p}) = p_y \in \mathbb{R} \\
\boldsymbol{F}_3 &\in \mathbb{R}^{d \times c} \quad \texttt{Proj}_3(\boldsymbol{p}) = p_z \in \mathbb{R}
\end{aligned}$$

Where w, h, and d are the spatial dimensions of the voxel grid we aim to represent, and c is a channel count. After interpolation, an element-wise (Hadamard) product $\odot$ is used to reduce interpolated latents $\boldsymbol{Z}_1$, $\boldsymbol{Z}_2$, and $\boldsymbol{Z}_3$ into the final latent $\boldsymbol{Z}$:

$$\texttt{Reduce}(\boldsymbol{Z}_1, \boldsymbol{Z}_2, \boldsymbol{Z}_3) = \boldsymbol{Z}_1 \odot \boldsymbol{Z}_2 \odot \boldsymbol{Z}_3 \tag{E.2}$$

E.2. Tri-plane architectures

Beginning in generative 3D [25, 33], several works have evaluated *tri-plane* architectures for decomposing latent 3D grids. The key idea of a tri-plane is to build feature grids along the XY, YZ, and XZ planes (Figure 2b), which are dramatically more compact than a full 3D tensor and conducive to generative architectures developed for 2D image synthesis. Using the notation described above, this can be concretized by setting $F = 3$ and defining three axis-aligned factors and projection functions:

$$\begin{aligned}
\boldsymbol{F}_1 &\in \mathbb{R}^{w \times h \times c} \quad \texttt{Proj}_1(\boldsymbol{p}) = (p_x, p_y) \in \mathbb{R}^2 \\
\boldsymbol{F}_2 &\in \mathbb{R}^{h \times d \times c} \quad \texttt{Proj}_2(\boldsymbol{p}) = (p_y, p_z) \in \mathbb{R}^2 \\
\boldsymbol{F}_3 &\in \mathbb{R}^{w \times d \times c} \quad \texttt{Proj}_3(\boldsymbol{p}) = (p_x, p_z) \in \mathbb{R}^2
\end{aligned}$$

As described in the general case above, projected coordinates are used to interpolate per-projection latent vectors $\boldsymbol{Z}_1$, $\boldsymbol{Z}_2$, and $\boldsymbol{Z}_3$ from the corresponding set of feature grids $\boldsymbol{F}_1$, $\boldsymbol{F}_2$, and $\boldsymbol{F}_3$, which are passed through a `Reduce` operation to produce the final latent vector $\boldsymbol{Z}$.

Several choices exist for `Reduce`. EG3D [25], which adapts a StyleGAN2 [21] architecture for 3D generation of faces and cats, uses element-wise summation:

$$\texttt{Reduce}(\boldsymbol{Z}_1, \boldsymbol{Z}_2, \boldsymbol{Z}_3) = \boldsymbol{Z}_1 + \boldsymbol{Z}_2 + \boldsymbol{Z}_3$$

Rodin [33], which adapts latent diffusion [28] for 3D generation of avatars, adopts concatenation:

$$\texttt{Reduce}(\boldsymbol{Z}_1, \boldsymbol{Z}_2, \boldsymbol{Z}_3) = \boldsymbol{Z}_1 \oplus \boldsymbol{Z}_2 \oplus \boldsymbol{Z}_3$$

Outside of generative models, K-Planes [35] demonstrates that a Hadamard product for reduction is advantageous when applied with both linear and MLP decoders. In TILTED, we adopt the K-Planes naming for tri-plane architectures due to the use of product-based reduction.

E.3. Vector-matrix pairs

Rather than building a representation using only matrix components, TensoRF [31] proposes a factorization of voxel grids using three vector-matrix (VM) pairs (Figure 2c). The corresponding factors and projection functions can be formalized as:

$$\begin{aligned}
\boldsymbol{F}_1 &\in \mathbb{R}^{w \times c} & \texttt{Proj}_1(\boldsymbol{p}) &= p_x \\
\boldsymbol{F}_2 &\in \mathbb{R}^{h \times d \times c} & \texttt{Proj}_2(\boldsymbol{p}) &= (p_y, p_z) \\
\boldsymbol{F}_3 &\in \mathbb{R}^{h \times c} & \texttt{Proj}_3(\boldsymbol{p}) &= p_y \\
\boldsymbol{F}_4 &\in \mathbb{R}^{w \times d \times c} & \texttt{Proj}_4(\boldsymbol{p}) &= (p_x, p_z) \\
\boldsymbol{F}_5 &\in \mathbb{R}^{h \times c} & \texttt{Proj}_5(\boldsymbol{p}) &= p_z \\
\boldsymbol{F}_6 &\in \mathbb{R}^{w \times h \times c} & \texttt{Proj}_6(\boldsymbol{p}) &= (p_x, p_y)
\end{aligned}$$

The result is 6 interpolated latent vectors $\boldsymbol{Z}_{1\ldots6}$. Components from each vector-matrix pair are multiplied to produce 3 vectors, which are then concatenated:

$$\texttt{Reduce}(\boldsymbol{Z}_{1\ldots6}) = \oplus_{i=1,3,5} \left[\boldsymbol{Z}_i \odot \boldsymbol{Z}_{i+1} \right]$$

After reduction, the latent $\boldsymbol{Z}$ is passed to an MLP decoder to regress quantities like radiance or signed distance.

E.4. Multi-resolution factors

The decomposition architectures presented in Sections E.1, E.2, and E.3 all assume that decompositions exist at only one resolution per scene. In practice, it can be advantageous to aggregate features at varying spatial resolutions [32, 35].

Adapting the notation above to the multi-resolution setting is straightforward. K-Planes, for example, runs all experiments at four resolutions: 64×64, 128×128, 256×256, and 512×512. Generalizing to R resolutions and per-resolution scale factor s_r, the process for interpolating multi-resolution K-Planes features can be written with our abstractions as:

$$\begin{aligned}
\boldsymbol{F}_{r,1} &\in \mathbb{R}^{w_r \times h_r \times c} & \texttt{Proj}_{r,1}(\boldsymbol{p}) &= (s_r p_x, s_r p_y) \\
\boldsymbol{F}_{r,2} &\in \mathbb{R}^{h_r \times d_r \times c} & \texttt{Proj}_{r,2}(\boldsymbol{p}) &= (s_r p_y, s_r p_z) \\
\boldsymbol{F}_{r,3} &\in \mathbb{R}^{w_r \times d_r \times c} & \texttt{Proj}_{r,3}(\boldsymbol{p}) &= (s_r p_x, s_r p_z)
\end{aligned}$$

for $r = 1 \ldots R$. For the `Reduce` operator, the Hadamard product is applied within each resolution, and concatenation is applied across resolutions:

$$\texttt{Reduce}(\{\boldsymbol{Z}_{r,i}\}_{r,i}) \quad = \oplus_{r=1\ldots R} \left[\boldsymbol{Z}_{r,1} \odot \boldsymbol{Z}_{r,2} \odot \boldsymbol{Z}_{r,3} \right]$$

TILTED applies this pattern to all 3D experiments.

F. Proofs for Section 3

We assume throughout these appendices that $n \geq 2$.

Notation. We write $\mathbb{R}$ for the reals, $\mathbb{Z}$ for the integers, and $\mathbb{N}$ for the positive integers. For positive integers m and n, we let $\mathbb{R}^m$ and $\mathbb{R}^{m\times n}$ denote the spaces of real-valued m-dimensional vectors and m-by-n matrices (resp.). We write $\boldsymbol{e}_i$, $\boldsymbol{e}_{ij}$, etc. to denote the elements of the canonical basis of these spaces, and $\mathbf{1}_m$ and $\mathbf{0}_{m,n}$ (etc.) to denote their all-ones and all-zeros elements (resp.). We write $\langle\,\cdot\,,\,\cdot\,\rangle$ and $\|\cdot\|_F$ to denote the euclidean inner product and associated norm of these spaces. We will write the ℓ^p norms $\|\boldsymbol{x}\|_p = (\sum_i |x_i|^p)^{1/p}$, with $\|\boldsymbol{x}\|_\infty = \max_i |x_i|$, of these spaces as either $\|\cdot\|_p$ or $\|\cdot\|_{\ell^p}$ depending on context. We will use the notation $\|\cdot\|$ to denote the operator norm (the largest singular value) on $m \times n$ matrices. If $\boldsymbol{A} \in \mathbb{R}^{m\times n}$, we write $\boldsymbol{A}^* \in \mathbb{R}^{n\times m}$ for its (conjugate) transpose. For matrices $\boldsymbol{A}$ and $\boldsymbol{B}$, we write $\boldsymbol{A} \otimes \boldsymbol{B}$ to denote their tensor product—if indices (i, j) index $\boldsymbol{A}$ and (k, l) index $\boldsymbol{B}$, we have $(\boldsymbol{A} \otimes \boldsymbol{B})_{ijkl} = (\boldsymbol{A})_{ij}(\boldsymbol{B})_{kl}$.

As a technical tool (in Section F.1), and as a mathematical abstraction (in Section F.2), we will frequently work with "continuum" images defined on the square $[-1, 1]^2 \subset \mathbb{R}^2$. By default, we will use "image coordinates" for $\boldsymbol{x} \in \mathbb{R}^2$ (in order to match the usual matrix-type indexing of discrete images), which corresponds in the canonical basis to the positively-oriented frame $[-\boldsymbol{e}_2, \boldsymbol{e}_1]$. We will formally write these coordinates as $\boldsymbol{x} = (s, t)$. For an image $X : \mathbb{R}^2 \to [0, 1]$ we will write $\|X\|_{L^p} = (\int_{\mathbb{R}^2} |X(\boldsymbol{x})|^p \,\mathrm{d}\boldsymbol{x})^{1/p}$ for the L^p norms, and $\|X\|_{L^\infty} = \sup_{\boldsymbol{x}\in\mathbb{R}^2} |X(\boldsymbol{x})|$ when X is bounded. The space $L^2(\mathbb{R}^d)$ is a Hilbert space; as for finite-dimensional vector spaces, we will write $\langle\,\cdot\,,\,\cdot\,\rangle_{L^2}$ for its associated inner product (which we take to be linear in its second argument), and if $\mathcal{T} : L^2 \to L^2$ is a bounded operator we will write $\|\mathcal{T}\|$ for its (operator) norm and $\mathcal{T}^*$ for its adjoint. Similarly, we will use notation defined above for matrix operations for its analogous application to L^2 functions (e.g., tensor products). If $\boldsymbol{\tau} : \mathbb{R}^2 \to \mathbb{R}^2$ is a continuous function (e.g., a rotation of the domain) and $X : \mathbb{R}^2 \to \mathbb{R}$ is an image, we write $X \circ \boldsymbol{\tau}$ for their composition (the "deformed image"). For sufficiently regular functions $f, g : \mathbb{R}^2 \to \mathbb{R}$, we define their convolution $(f * g)(\boldsymbol{x}) = \int_{\mathbb{R}^2} f(\boldsymbol{x}')g(\boldsymbol{x} - \boldsymbol{x}') \,\mathrm{d}\boldsymbol{x}'$; this operation is symmetric and defines an element of L^p when (say) f is in L^1 and g is in L^p. We will use $\mathbb{1}_A$ to denote the indicator function associated to an event A in a probability space; typically A will be a subset of $\mathbb{R}^2$ (e.g., describing a continuous image) or a discrete set (e.g., describing the Kronecker delta $\mathbb{1}_{i=j}$ in summations).

Just as with discrete images, which can either be thought of as a function on the discrete grid $\{0, \dots, m-1\}\times\{0, \dots, n-1\}$, representing sampled intensity values, or a matrix (i.e., a finite-dimensional operator) that aggregates those values, "continuum images" can also be thought of as either functions or operators; if $f \in L^2(\mathbb{R}^2)$, we will write $\mathcal{T}_f : L^2(\mathbb{R}) \to L^2(\mathbb{R})$ for the "Fredholm operator" associated to an L^2 function f, defined by $\mathcal{T}_f[g] = \int_{\mathbb{R}} f(\,\cdot\,, x)g(x) \,\mathrm{d}x$. If $\mathcal{T} : L^2(\mathbb{R}) \to L^2(\mathbb{R})$ is bounded, we denote its Hilbert-Schmidt norm by $\|\mathcal{T}\|_{\text{HS}} = (\sum_{n\in\mathbb{N}} \|\mathcal{T}u_n\|^2_{L^2(\mathbb{R})})^{1/2}$, where $(u_n)_{n\in\mathbb{N}}$ is any orthonormal basis of $L^2(\mathbb{R})$; when $\mathcal{T}_f$ is a Fredholm operator, we have $\|\mathcal{T}_f\|_{\text{HS}} = \|f\|_{L^2(\mathbb{R}^2)}$, analogous to the Frobenius norm of a matrix. We will exploit this correspondence in the sequel, often without mentioning it, to identify a function $f \in L^2(\mathbb{R}^2)$ with its Fredholm operator $\mathcal{T}_f$ when convenient (c.f. [12, §B]); for example, for $f \in L^2(\mathbb{R})$ we will write $ff^* : L^2(\mathbb{R}) \to L^2(\mathbb{R})$ to denote its induced Fredholm operator, which satisfies $ff^*[g] = f\langle f, g\rangle_{L^2(\mathbb{R})}$, and we will identify it with its $L^2(\mathbb{R}^2)$ representative satisfying $ff^*(s, t) = f(s)f(t)$. Consult the first few paragraphs in Section F.1 for specialized notation used in low-rank approximation proofs, and the proof of Lemma F.17 for notation used in proofs that require harmonic analysis.

Problem setup. We analyze a simple model problem that captures the improved efficiency of TILTED compared to competing approaches for compactly representing non-axis-aligned scenes. Consider the following class of two-dimensional greyscale images: let $m, n \in \mathbb{N}$ denote the image height and width, write $\boldsymbol{c} = [\frac{m-1}{2}, \frac{n-1}{2}]^*$ for the image center (we use zero-indexing), and define a centered square template by

$$(\boldsymbol{X}_\natural)_{ij} = \begin{cases} 1 & \|[i, j]^* - \boldsymbol{c}\|_\infty \le \alpha \min\{c_0, c_1\} \\ 0 & \text{otherwise,} \end{cases} \tag{F.1}$$

where $0 < \alpha < 1$ controls the size of the square; we are interested in $\alpha < 1/\sqrt{2}$, for a square that takes up a constant fraction of the image pixels. We consider a rotational motion model for the square template $\boldsymbol{X}_\natural$: for a parameter $\nu \in [0, 2\pi)$ corresponding to the rotation about the image center $\boldsymbol{c}$, let $\boldsymbol{\tau}_\nu : \mathbb{R}^2 \to \mathbb{R}^2$ denote the (continuum) transformation corresponding to

$$\begin{bmatrix} s \\ t \end{bmatrix} \mapsto \begin{bmatrix} \cos\nu & -\sin\nu \\ \sin\nu & \cos\nu \end{bmatrix} \left(\begin{bmatrix} s \\ t \end{bmatrix} - \boldsymbol{c} \right) + \boldsymbol{c}, \tag{F.2}$$

and consider the class of observations

$$\mathfrak{S} = \left\{ \boldsymbol{X} \in \mathbb{R}^{m\times n} \;\middle|\; X_{ij} = \begin{cases} 1 & \|\boldsymbol{\tau}_{-\nu}(i, j) - \boldsymbol{c}\|_\infty \le \alpha \min\{c_0, c_1\} \\ 0 & \text{otherwise} \end{cases} \right\}. \tag{F.3}$$

In our lower bounds on low-rank compression in Section F.1, we will work with a "directly-sampled" observation following the model (F.3). In Section F.2, we will work in a continuum idealization where it is more convenient to describe the observations in a shifted coordinate system, which we now describe.

In our proofs, we will work in a shifted coordinate system so that the center of the square (F.1) lies at the origin of the coordinate system . In particular, in these appendices we consider the image grid $\{0, 1, \dots, m-1\} \times \{0, 1, \dots, n-1\} - \boldsymbol{c}$, corresponding to the grid

$$G_{\boldsymbol{c}} = \{(i,j) \mid i \in \{-(m-1)/2, \dots, (m-1)/2\}, j \in \{-(n-1)/2, \dots, (n-1)/2\}\}.$$

We will often index vectors and matrices by their coordinates in $G_{\boldsymbol{c}}$ and its derived grids, rather than in the standard image grid, due to the straightforward one-to-one correspondence between grids. Without loss of generality, we will assume that $m \leq n$. Let us then note that in $G_{\boldsymbol{c}}$ coordinates, (F.1) admits the equivalent rank-one expression

$$\boldsymbol{X}_\natural = \boldsymbol{u}_\natural \boldsymbol{v}_\natural^*, \quad (\boldsymbol{u}_\natural)_i = \begin{cases} 1 & |i| \leq \frac{\alpha}{2}(m-1) \\ 0 & \text{otherwise,} \end{cases}, \quad (\boldsymbol{v}_\natural)_j = \begin{cases} 1 & |j| \leq \frac{\alpha}{2}(m-1) \\ 0 & \text{otherwise,} \end{cases}. \tag{F.4}$$

We will require, roughly, that $0 < \alpha < \frac{1}{\sqrt{2}}$, so that there are no boundary issues with rotated versions of the template (F.4).

The template definition (F.4) implies that as the image size m, n become large, $\boldsymbol{X}_\natural$ samples the same fixed continuum template $X_\natural : [-1,1] \to \{0,1\}$ defined by

$$X_\natural(s,t) = \mathbb{1}_{|s| \leq \alpha, |t| \leq \alpha}. \tag{F.5}$$

To make this correspondence, it is necessary to scale the grid $G_{\boldsymbol{c}}$ by the factor $2/(m-1)$: this corresponds to the grid

$$G = \left\{ (i,j) \;\middle|\; i \in \left\{-1, -1 + \frac{2}{m-1}, \dots, 1 - \frac{2}{m-1}, 1\right\}, j \in \left\{-\frac{n-1}{m-1}, \dots, \frac{n-1}{m-1}\right\} \right\}. \tag{F.6}$$

It is then evident that if $(i,j) \in G$, one has $(\boldsymbol{X}_\natural)_{ij} = X_\natural(i,j)$.

The possible complication that one may have rectangular images with $n > m$ is actually not essential—to see this, note that we always have the block structure

$$\boldsymbol{X}_\natural = \begin{bmatrix} \boldsymbol{0} & \bar{\boldsymbol{X}}_\natural & \boldsymbol{0}', \end{bmatrix}$$

where $\bar{\boldsymbol{X}}_\natural$ follows the definition (F.4), but with $m = n$, and $\boldsymbol{0}$ and $\boldsymbol{0}'$ are zero matrices of appropriate sizes. This shows that $\boldsymbol{X}_\natural$ and $\bar{\boldsymbol{X}}_\natural$ have the same nonzero singular values, the same left singular vectors, and right singular vectors that are in one-to-one correspondence (simply prepend and append the appropriate number of zeros to the singular vectors of $\bar{\boldsymbol{X}}_\natural$). This implies that in our proofs for the SVD approach in Section F.1, we may assume that $m = n$ without any loss of generality.

F.1. Proofs for Theorem 1

As mentioned previously, without loss of generality we assume $m = n$ in this section.

Problem setting. We study the special case of $\nu_\natural = \pi/4$, so that the observation

$$(\boldsymbol{X})_{ij} = \mathbb{1}_{\|(\boldsymbol{\tau}_{\pi/4})_{ij}\|_\infty \leq \alpha}$$

corresponds to a "diamond". This case makes the rank of the transformed image as large as possible.

Continuum surrogate. Our analysis will proceed by relating the singular value decomposition of $\boldsymbol{X}_\natural \circ \boldsymbol{\tau}_{\nu_\natural}$ to the spectrum of an 'infinite resolution' surrogate X, defined as

$$X(s,t) = X_\natural(s \cos \nu_\natural + t \sin \nu_\natural, -s \sin \nu_\natural + t \cos \nu_\natural).$$

Whereas $\operatorname{supp} X_\natural = [-\alpha, \alpha]^2$, we have $\operatorname{supp} X = [-\sqrt{2}\alpha, \sqrt{2}\alpha]^2$. The 'infinite resolution' analogue of taking the singular value decomposition of an image is the Schmidt decomposition (c.f. [8]) of the image's associated Fredholm operator: define $\mathcal{T}_X : L^2([-1,+1]) \to L^2([-1,+1])$ by

$$\mathcal{T}_X[f](s) = \int_{[-1,+1]} X(s,t) f(t) \, \mathrm{d}t,$$

and note by the geometry of the diamond X that

$$\mathcal{T}_X[f](s) = \int_{-(\sqrt{2}\alpha - |s|)}^{\sqrt{2}\alpha - |s|} f(t)\,\mathrm{d}t, \tag{F.7}$$

so that in particular $\mathcal{T}_X$ is self-adjoint and Hilbert-Schmidt (hence compact). The spectral theorem for compact operators on a Hilbert space [12] then implies that $\mathcal{T}_X$ diagonalizes in an orthonormal basis of eigenfunctions $(e_k)_{k\in\mathbb{N}} \subset L^2([-1,+1])$ with corresponding eigenvalues $(\lambda_k)_{k\in\mathbb{N}} \subset \mathbb{R}$:

$$\mathcal{T}_X = \sum_{k\in\mathbb{N}} \lambda_k e_k e_k^*, \tag{F.8}$$

where the equality must be interpreted in the sense of $L^2 \to L^2$. We will derive a closed-form expression for (F.8) for the diamond (Lemma F.1), and use a truncation and discretization of it as an approximate diagonalization of the discrete diamond $\boldsymbol{X}$.

Approximation guarantees with the SVD. The use of an infinite-dimensional surrogate to analyze $\boldsymbol{X}$ requires the instantiation of some approximation machinery. We quantify reconstruction performance in terms of squared error. For any matrix $\boldsymbol{M} \in \mathbb{R}^{n\times n}$, we write $\sigma_1(\boldsymbol{M}) \geq \sigma_2(\boldsymbol{M}) \geq \cdots \geq \sigma_n(\boldsymbol{M}) \geq 0$ for its singular values. The singular value decomposition asserts that for any $\boldsymbol{M}$, there exist orthogonal matrices $\boldsymbol{U}(\boldsymbol{M})$ and $\boldsymbol{V}(\boldsymbol{M})$ such that

$$\boldsymbol{M} = \boldsymbol{U} \underbrace{\begin{bmatrix} \sigma_1(\boldsymbol{M}) & & \\ & \ddots & \\ & & \sigma_n(\boldsymbol{M}) \end{bmatrix}}_{\boldsymbol{\Sigma}} \boldsymbol{V}^*.$$

We recall that $\|\boldsymbol{M}\|_{\mathrm{F}}^2 = \sum_{i=1}^n \sigma_i^2(\boldsymbol{M})$. The "rank-$k$" SVD approximation to $\boldsymbol{M}$ is defined as[1]

$$\mathrm{SVD}_k(\boldsymbol{M}) = \boldsymbol{U} \begin{bmatrix} \sigma_1(\boldsymbol{M}) & & & \\ & \ddots & & \\ & & \sigma_k(\boldsymbol{M}) & \\ & & & \boldsymbol{0}_{n-k,n-k} \end{bmatrix} \boldsymbol{V}^*.$$

Following [7], we write $\|\boldsymbol{M}\|_{(k)}^{(p)} = \left(\sum_{i=1}^k \sigma_i^p(\boldsymbol{M})\right)^{1/p}$ for the Ky Fan p-norms of a matrix $\boldsymbol{M}$. These are indeed norms in the mathematical sense (e.g., [7, §IV.2, eqn. IV.47]). From the celebrated Eckart-Young-Mirsky theorem [1, 2], we have

$$\inf_{\mathrm{rank}(\boldsymbol{M})\leq k} \|\boldsymbol{M} - \boldsymbol{X}\|_{\mathrm{F}}^2 = \sum_{i=k+1}^n \sigma_i^2(\boldsymbol{X}) = \|\boldsymbol{X}\|_{\mathrm{F}}^2 - \left(\|\boldsymbol{X}\|_{(k)}^{(2)}\right)^2,$$

and and it is evident that $\boldsymbol{M} = \mathrm{SVD}_k(\boldsymbol{X})$ achieves the infimum in this formula: that is,

$$\|\mathrm{SVD}_k(\boldsymbol{X}) - \boldsymbol{X}\|_{\mathrm{F}}^2 = \|\boldsymbol{X}\|_{\mathrm{F}}^2 - \left(\|\boldsymbol{X}\|_{(k)}^{(2)}\right)^2.$$

It follows that we can obtain lower bounds on the approximation error of SVD-based compression of $\boldsymbol{X}$ via upper bounds on the Ky Fan 2-norms of $\boldsymbol{X}$.

For any $\boldsymbol{\Xi} \in \mathbb{R}^{n\times n}$, we have from the triangle inequality

$$\begin{aligned} \|\boldsymbol{X}\|_{(k)}^{(2)} &\leq \|\boldsymbol{\Xi}\|_{(k)}^{(2)} + \|\boldsymbol{\Xi} - \boldsymbol{X}\|_{(k)}^{(2)} \\ &\leq \|\boldsymbol{\Xi}\|_{(k)}^{(2)} + \|\boldsymbol{\Xi} - \boldsymbol{X}\|_{\mathrm{F}}, \end{aligned} \tag{F.9}$$

[1] The "scare quotes" are to draw attention to the fact that if $\boldsymbol{M}$ has rank strictly less than k, this approximation is not actually rank k—its rank is no larger than $\mathrm{rank}(\boldsymbol{M})$.

where the second inequality simply worst-cases over all n singular values of the residual. (F.9) is the basis of our approximation argument: we will choose $\boldsymbol{\Xi}$ as a matrix whose spectral decay is known, and which gives a good approximation to the actual diamond matrix $\boldsymbol{X}$. In particular, we will consider a family of approximations $\boldsymbol{\Xi}_m$, with $m \in \mathbb{N}$, defined as

$$(\boldsymbol{\Xi}_m)_{ij} = \sum_{l=1}^{m} \lambda_l g_l(i) g_l(j), \tag{F.10}$$

with coordinates $(i,j) \in G$ and with notation as defined in Lemma F.1. We discuss the sources of error in these approximations momentarily; let us first introduce additional notation to write these matrices more compactly. Define $\boldsymbol{U}_m \in \mathbb{R}^{n \times m}$ by

$$(\boldsymbol{U}_m)_{ij} = g_j(i); \quad i \in \{k \mid \exists l : (k,l) \in G\}, \ \ j \in [n],$$

and let $\boldsymbol{\Lambda}_m \in \mathbb{R}^{m \times m}$ be a diagonal matrix with λ_l on its l-th diagonal entry. Then

$$\boldsymbol{\Xi}_m = \boldsymbol{U}_m \boldsymbol{\Lambda}_m \boldsymbol{U}_m^*. \tag{F.11}$$

For technical reasons, we will need to consider a further level of approximation induced by smoothing the nonsmooth square pattern $X_\natural$. For $\sigma^2 > 0$, we write $\varphi_{\sigma^2}(t) = 1/\sqrt{2\pi\sigma^2}\exp(-\frac{1}{2\sigma^2}t^2)$ for the one-dimensional standard gaussian, and $m_{\sigma^2} = \varphi_{\sigma^2}^{\otimes 2}$ for its two-dimensional analogue. Let $f * g$ denote the convolution of $L^2(\mathbb{R}^2)$ signals f and g. Then define a smoothed family of approximations

$$(\tilde{\boldsymbol{\Xi}}_m)_{ij} = \sum_{l=1}^{m} \lambda_l (\varphi_{\sigma^2} * g_l)(i)(\varphi_{\sigma^2} * g_l)(j). \tag{F.12}$$

As above, let $\tilde{\boldsymbol{\Xi}}_m$ denote the matrix representation of this construction:

$$\tilde{\boldsymbol{\Xi}}_m = \tilde{\boldsymbol{U}}_m \boldsymbol{\Lambda}_m \tilde{\boldsymbol{U}}_m^*. \tag{F.13}$$

Relative to the continuum diamond X, there are three main sources of error in the approximations (F.12). The parameter m controls a truncation of the infinite series of eigenfunctions that defines $\mathcal{T}_X$, and the grid resolution (proportional to n) controls a discretization error relative to the continuum image X. In addition, the smoothing scale σ^2 controls a further error, since the smoothed eigenfunctions do not coincide with eigenfunctions of the 'smoothed operator'. These three parameters are in tension—choosing m larger recovers more terms in the series defining $\mathcal{T}_X$, but when the grid resolution is fixed at $2/(n-1)$, the fact that the eigenfunctions g_l become more and more oscillatory at larger values of l suggests a larger and larger discretization error, and a need for a smaller and smaller smoothing scale σ^2 to avoid destroying the spectral structure of the eigenfunctions g_l. We will choose these parameters in tandem with the SVD rank k in (F.9) in order to guarantee as strong of a lower bound on the approximation error as possible.

Main result. Our main result is an inapproximability result for sublinear low-rank approximations to $\boldsymbol{X}_\natural \circ \boldsymbol{\tau}_\nu$, up to a threshold.

Theorem F.1. *There are absolute constants $c, C, C' > 0$ such that the following holds. Let $\nu = \pi/4$ and $\alpha = 1/\sqrt{2}$, and consider the observation*

$$(\boldsymbol{X}_\nu)_{ij} = \mathbb{1}_{\|(\boldsymbol{\tau}_\nu)_{ij}\|_\infty \le \alpha}.$$

For every $n \ge \max\{C, C'k^6\}$, one has for every $\hat{\boldsymbol{X}} \in \mathbb{R}^{n \times n}$ with rank no larger than k

$$\frac{1}{n^2}\left\|\hat{\boldsymbol{X}} - \boldsymbol{X}_\nu\right\|_{\mathrm{F}}^2 \ge \frac{c}{1+k}.$$

Proof. We instantiate the argument discussed in the previous paragraph, culminating in (F.9). Below, we will occasionally not calculate precise constants for simplicity, and similarly we will fix $\alpha = 1/\sqrt{2}$, allowing us to treat it as an absolute constant. Put

$$\bar{X} = \varphi_{\sigma^2}^{\otimes 2} * X_\nu$$

for the smoothed observation (note that $(\boldsymbol{X}_\nu)_{ij} = X_\nu(i,j)$ for $(i,j) \in G$), let $\bar{G} = (-1,-1) + \frac{2}{n-1}\mathbb{Z}^2$ denote the infinitely-extended grid G defined in (F.6), and let $(\bar{\boldsymbol{X}})_{ij} = \bar{X}(i,j)$ for $(i,j) \in \bar{G}$. The inclusion $G \subset \bar{G}$ means that we can naturally

think of $\bar{\boldsymbol{X}}$ as a matrix indexed by G as well (via restriction), and we will write $\|\cdot\|_{\ell^2(G)}$ and $\|\cdot\|_{\ell^2(\bar{G})}$ to denote the respective norms. Observe that, by linearity of the convolution operation and Lemma F.1, we have for $(i,j) \in \bar{G}$

$$\begin{aligned}(\bar{\boldsymbol{X}})_{ij} &= (\varphi_{\sigma^2}^{\otimes 2} * X_\nu)(i,j) \\ &= \sum_{l=1}^{\infty} \lambda_l (\varphi_{\sigma^2} * g_l)(i)(\varphi_{\sigma^2} * g_l)(j) \\ &= (\tilde{\boldsymbol{\Xi}})_{ij} + \left(\underbrace{\sum_{l=m+1}^{\infty} \lambda_l (\varphi_{\sigma^2} * g_l)(\varphi_{\sigma^2} * g_l)^*}_{\Delta_{\text{tail}}} \right)(i,j). \end{aligned} \tag{F.14}$$

Let $\hat{\boldsymbol{X}}$ be any approximation to $\boldsymbol{X}_\nu$ with rank at most k. For technical convenience, we want to compare $\ell^2(G)$ norms to $\ell^2(\bar{G})$ norms—note that these are distinct when we consider our smoothed approximation $\bar{\boldsymbol{X}}$, because convolution with the mollifier enlarges the support to be outside of $[-1,1]^2$. We extend $\hat{\boldsymbol{X}}$ and $\boldsymbol{X}_\nu$ to all of $\bar{G}$ by zero-padding, and note that

$$\left\|\hat{\boldsymbol{X}} - \boldsymbol{X}_\nu\right\|_{\text{F}} = \left\|\hat{\boldsymbol{X}} - \boldsymbol{X}_\nu\right\|_{\ell^2(G)} = \left\|\hat{\boldsymbol{X}} - \boldsymbol{X}_\nu\right\|_{\ell^2(\bar{G})}.$$

By the triangle inequality, we have

$$\left\|\hat{\boldsymbol{X}} - \boldsymbol{X}_\nu\right\|_{\ell^2(G)} \geq \left\|\hat{\boldsymbol{X}} - \bar{\boldsymbol{X}}\right\|_{\ell^2(G)} - \left\|\bar{\boldsymbol{X}} - \boldsymbol{X}_\nu\right\|_{\ell^2(G)},$$

We can thus apply the EYM theorem to obtain

$$\left\|\hat{\boldsymbol{X}} - \bar{\boldsymbol{X}}\right\|_{\ell^2(G)} \geq \sqrt{\|\bar{\boldsymbol{X}}\|_{\ell^2(G)}^2 - \left(\|\bar{\boldsymbol{X}}\|_{(k)}^{(2)}\right)^2}. \tag{F.15}$$

Notice that, by (F.14) and the fact that the Ky Fan 2-norms are mathematically norms, we have

$$\begin{aligned}\left(\|\bar{\boldsymbol{X}}\|_{(k)}^{(2)}\right)^2 &\leq \left(\left\|\tilde{\boldsymbol{\Xi}}\right\|_{(k)}^{(2)} + \|\boldsymbol{\Delta}_{\text{tail}}\|_{(k)}^{(2)}\right)^2 \\ &= \left(\left\|\tilde{\boldsymbol{\Xi}}\right\|_{(k)}^{(2)}\right)^2 + \left(\|\boldsymbol{\Delta}_{\text{tail}}\|_{(k)}^{(2)}\right)^2 + 2\left\|\tilde{\boldsymbol{\Xi}}\right\|_{(k)}^{(2)} \|\boldsymbol{\Delta}_{\text{tail}}\|_{(k)}^{(2)} \\ &\leq \left(\left\|\tilde{\boldsymbol{\Xi}}\right\|_{(k)}^{(2)}\right)^2 + \|\boldsymbol{\Delta}_{\text{tail}}\|_{\ell^2(G)}^2 + 2\left\|\tilde{\boldsymbol{\Xi}}\right\|_{(k)}^{(2)} \|\boldsymbol{\Delta}_{\text{tail}}\|_{\ell^2(G)} \\ &\leq \left(\left\|\tilde{\boldsymbol{\Xi}}\right\|_{(k)}^{(2)}\right)^2 + \|\boldsymbol{\Delta}_{\text{tail}}\|_{\ell^2(\bar{G})}^2 + 2\left\|\tilde{\boldsymbol{\Xi}}\right\|_{(k)}^{(2)} \|\boldsymbol{\Delta}_{\text{tail}}\|_{\ell^2(\bar{G})}. \end{aligned}$$

Moreover, by Lemmas F.2 and F.3, we have

$$\begin{aligned}\left(\left\|\tilde{\boldsymbol{\Xi}}_m\right\|_{(k)}^{(2)}\right)^2 \leq \frac{n^2}{4}\left(4\alpha^2 - \frac{16\alpha^2}{\pi^2}\frac{1}{2\min\{m,k\}+1}\right) &+ Cn(m(1+\log m)^{1/2} + n\sigma^2 m^2) \\ &+ C'(m^2(1+\log m) + n^2\sigma^4 m^4).\end{aligned}$$

Meanwhile, by (F.14), we have that Δ_{tail} is in $L^1(\mathbb{R}^2)$, and its L^1 norm is no larger than that of $\bar{X}$. Applying Lemma F.17 thus implies

$$\|\boldsymbol{\Delta}_{\text{tail}}\|_{\ell^2(\bar{G})}^2 \leq \frac{n^2}{4}\|\Delta_{\text{tail}}\|_{L^2}^2 + \frac{C}{\sigma^4}(1+n\sigma).$$

By Young's inequality, we have that $\|\Delta_{\text{tail}}\|_{L^2}$ is less than the corresponding tail sum without smoothing. Now notice that, by orthogonality,

$$\left\| \sum_{l=m+1}^{\infty} \lambda_l g_l g_l^* \right\|_{L^2}^2 = \sum_{l=m+1}^{\infty} \lambda_l^2 \leq \frac{32\alpha^2}{\pi^2} \frac{1}{2m+1},$$

following the arguments in the proof of Lemma F.3 (the last estimate assumes $m \geq 1$). Thus

$$\|\boldsymbol{\Delta}_{\text{tail}}\|_{\ell^2(\bar{G})}^2 \leq \frac{32\alpha^2 n^2}{4\pi^2} \frac{1}{2m+1} + \frac{1}{\sigma^4}(1+n\sigma).$$

Combining these estimates, we have

$$\left(\|\bar{\boldsymbol{X}}\|_{(k)}^{(2)}\right)^2 \leq n^2\alpha^2 + \frac{4n^2}{\pi^2}\frac{1}{2m+1} - \frac{2n^2/\pi^2}{2\min\{m,k\}+1} + Cn(m(1+\log m)^{1/2} + n\sigma^2 m^2) + C'(m^2(1+\log m) + n^2\sigma^4 m^4)$$
$$+ C''(1+n\sigma)/\sigma^4 + C'''\left(n + \sqrt{nm\log^{1/2} m} + n\sigma m + m\sqrt{\log m} + n\sigma^2 m^2\right)\left(\frac{n}{\sqrt{m}} + \sqrt{\frac{1+n\sigma}{\sigma^4}}\right).$$

To simplify the residual term, we will choose $m = n^{1/6}$ and $\sigma = m^{-3/2}$. Evaluating the residual in the previous expression shows that for n sufficiently large, there is an absolute constant $C > 0$ such that

$$\left(\|\bar{\boldsymbol{X}}\|_{(k)}^{(2)}\right)^2 \leq n^2\alpha^2 + \frac{4n^2}{\pi^2}\frac{1}{2m+1} - \frac{2n^2/\pi^2}{2\min\{m,k\}+1} + Cn^{23/12}.$$

Similarly, when $m \geq Ck$ for a sufficiently large constant C, this bound is upper bounded by

$$\left(\|\bar{\boldsymbol{X}}\|_{(k)}^{(2)}\right)^2 \leq n^2\alpha^2 - \frac{Cn^2}{k+1} + C'n^{23/12}.$$

Now, plugging this estimate into (F.15), we have

$$\left\|\hat{\boldsymbol{X}} - \boldsymbol{X}_\nu\right\|_{\ell^2(G)} \geq \sqrt{\|\bar{\boldsymbol{X}}\|_{\ell^2(G)}^2 - n^2\alpha^2 + \frac{Cn^2}{1+k} - C'n^{23/12}} - \left\|\bar{\boldsymbol{X}} - \boldsymbol{X}_\nu\right\|_{\ell^2(G)}.$$

We just need to estimate the remaining error terms. By Lemma F.4, we have for $m \geq 2^{2/3}$

$$\left\|\bar{\boldsymbol{X}} - \boldsymbol{X}_\nu\right\|_{\ell^2(G)}^2 \leq \frac{n^2\sigma^8}{\pi^2} + \frac{2n}{\pi} + \frac{n^2\sqrt{48\sigma^2\log(1/\sigma)}}{\pi}.$$

We have chosen $\sigma = n^{-1/4}$, which makes the residuals in this expression of order $n^{3/2}$: for sufficiently large n,

$$\left\|\bar{\boldsymbol{X}} - \boldsymbol{X}_\nu\right\|_{\ell^2(G)}^2 \leq Cn^{3/2}$$

for an absolute constant $C > 0$. Similarly, by this last bound and Lemma F.5 together with the triangle inequality, we have for n sufficiently large

$$\begin{aligned}\|\bar{\boldsymbol{X}}\|_{\ell^2(G)}^2 &\geq \left(\|\boldsymbol{X}_\nu\|_{\ell^2(G)} - \left\|\bar{\boldsymbol{X}} - \boldsymbol{X}_\nu\right\|_{\ell^2(G)}\right)^2 \\ &= \|\boldsymbol{X}_\nu\|_{\ell^2(G)}^2 - 2\|\boldsymbol{X}_\nu\|_{\ell^2(G)}\left\|\bar{\boldsymbol{X}} - \boldsymbol{X}_\nu\right\|_{\ell^2(G)} \\ &\geq n^2\alpha^2 - 5n - Cn^{7/4}\end{aligned}$$

where we use the trivial upper bound $\|\boldsymbol{X}_\nu\|_{\ell^2(G)} \leq n$. Plugging into our previous EYM estimate and noticing that the previous residual dominates, we have for n sufficiently large

$$\left\|\hat{\boldsymbol{X}} - \boldsymbol{X}_\nu\right\|_{\ell^2(G)} \geq \sqrt{\frac{Cn^2}{1+k} - C'n^{23/12}} - C'n^{3/4}.$$

We have chosen $\sigma = n^{-1/4}$, which makes the residuals of order $n^{3/2}$, which is lower order than the existing residual $n^{23/12}$. Choosing n sufficiently large, we can find absolute constants $c, C, C' > 0$ such that

$$\left\|\hat{\boldsymbol{X}} - \boldsymbol{X}_\nu\right\|_{\ell^2(G)} \geq \sqrt{\frac{cn^2}{1+k} - Cn^{23/12}} - C'n^{3/4}.$$

For the RMSE, this gives

$$\frac{1}{n^2}\left\|\hat{\boldsymbol{X}} - \boldsymbol{X}_\nu\right\|^2_{\ell^2(G)} \geq \frac{c}{1+k} - Cn^{-1/12} - 2C'n^{-1/4}\sqrt{\frac{c}{1+k} - Cn^{-1/12}}.$$

Because we have $k \leq Cm = Cn^{1/6}$, we have for large n

$$\frac{1}{n^2}\left\|\hat{\boldsymbol{X}} - \boldsymbol{X}_\nu\right\|^2_{\ell^2(G)} \geq \frac{c}{1+k} - C'\frac{1}{n^{1/4}\sqrt{k}},$$

and when $k \leq Cn^{1/2}$ for a certain absolute constant C, it follows

$$\frac{1}{n^2}\left\|\hat{\boldsymbol{X}} - \boldsymbol{X}_\nu\right\|^2_{\ell^2(G)} \geq \frac{c}{1+k}$$

for a sufficiently small absolute constant c. Because $k \leq Cm$ and $m = n^{1/6}$, this condition is satisfied.

□

Remark F.1. Theorem F.1 asserts lower bounds up to a threshold $k \lesssim n^{1/6}$. Based on empirical evidence and certain key residuals in our proofs, we believe it should be possible to assert the same lower bound up to scalings $k \lesssim n/\log^c(n)$, for some $c > 0$, although our arguments are insufficient to this task. The main technical issue we contend with in the proof of Theorem F.1 is the nonsmoothness of the underlying image $X_\natural$, which in our case necessitates the use of somewhat technical smoothing arguments. Some lemmas that we develop to this end, especially Lemmas F.17 and F.18, are suboptimal, and improvements would improve the rates. The crux of our argument should be applicable to templates $X_\natural$ that have better regularity without having to go through smoothing arguments, which should yield improved rates.

F.1.1 Supporting Results

Lemma F.1. *Define a sequence*

$$\lambda_k = (-1)^{k-1}\frac{4\sqrt{2}\alpha}{\pi(2k-1)}, \quad k = 1, 2, \ldots, \tag{F.16}$$

and functions $g_k : [-1, 1] \to \mathbb{R}$ *by*

$$g_k(s) = \begin{cases} \frac{1}{\sqrt{\alpha\sqrt{2}}}\cos\left(\frac{\pi}{2\sqrt{2}\alpha}(2k-1)s\right) & |s| \leq \sqrt{2}\alpha \\ 0 & \text{otherwise,} \end{cases} \quad k = 1, 2, \ldots \tag{F.17}$$

Then the functions g_k *form an orthonormal basis for the range of the (compact, self-adjoint) operator* $\mathcal{T}_X$*, and we have the decomposition*

$$\mathcal{T}_X = \sum_{k\in\mathbb{N}} \lambda_k g_k g_k^*.$$

Proof. We take the formula (F.7) as our starting point. Because of the spectral theorem for self-adjoint compact operators on a Hilbert space, we have the decomposition (F.8) for $\mathcal{T}_X$. Our approach will be to study the eigenvalue equation

$$\mathcal{T}_X[g] = \lambda g, \quad \lambda \neq 0, g \neq 0, \tag{F.18}$$

and to produce a large enough family of solutions (λ, g) to this equation that we can assert that we have produced the eigenvalues and eigenfunctions asserted by the spectral theorem in (F.8). To begin, we make several preliminary observations about solutions to the eigenvalue equation (F.18). First, we note from (F.7) and the change of variables formula that

$$\mathcal{T}_X[f](\sqrt{2}\alpha s) = \sqrt{2}\alpha\int_{-(1-|s|)}^{1-|s|} f(\sqrt{2}\alpha t)\,\mathrm{d}t,$$

so that, if for $\varepsilon > 0$ we write $\mathcal{S}_\varepsilon[g](u) = g(\varepsilon u)$ as the dilation operator (which satisfies $\mathcal{S}_\varepsilon^{-1} = \mathcal{S}_{\varepsilon^{-1}}$), we have

$$\mathcal{T}_X = \mathcal{S}_{\sqrt{2}\alpha} \bar{\mathcal{T}}_X \mathcal{S}_{\sqrt{2}\alpha}^{-1}, \tag{F.19}$$

where $\bar{\mathcal{T}}_X : L^2([-1,1]) \to L^2([-1,1])$ is defined as

$$\bar{\mathcal{T}}_X[f](s) = \sqrt{2}\alpha \int_{-(1-|s|)}^{1-|s|} f(t)\,\mathrm{d}t.$$

In particular, $\mathcal{T}_X$ is similar to the operator $\bar{\mathcal{T}}_X$. We therefore focus our analysis on $\bar{\mathcal{T}}_X$ below. Next, note that by the Schwarz inequality, we have

$$\begin{aligned}\left|\bar{\mathcal{T}}_X[f](s)\right| &\le \sqrt{2}\alpha \|f\|_{L^2} \|\mathbb{1}_{[-(1-|s|),1-|s|]}\|_{L^2} \\ &= 4\alpha \|f\|_{L^2} \sqrt{1-|s|}.\end{aligned}$$

In particular, we have $\bar{\mathcal{T}}_X[f](\pm 1) = 0$ for any $f \in L^2$. Thus, if f is moreover a solution to (F.18), it is necessary that $f(\pm 1) = 0$, giving us boundary conditions for the eigenvalue equation. Similarly, the formula (F.7) shows that $\bar{\mathcal{T}}_X[f](s) = \bar{\mathcal{T}}_X[f](-s)$ for any $f \in L^2$, so any f solving (F.18) also satisfies even symmetry.

We proceed with a standard bootstrapping argument—we start by seeking only solutions to (F.18) that are infinitely differentiable. For any $|s| > 0$, differentiating (F.7) gives the equivalent boundary value problem

$$\lambda g'(s) = -\sqrt{2}\alpha \,\mathrm{sign}(s)\left(g(1-|s|) - g(-(1-|s|))\right), \quad g(\pm 1) = 0$$

for the eigenvalue equation (F.18). By even symmetry of g, this is equivalent to the problem

$$\lambda g'(s) = -2\sqrt{2}\alpha\, g(1-s), \quad g(1) = 0, \;\; g'(0) = 0$$

with $g \in C^\infty([0,1])$. Differentiating once more to eliminate the 'space reversal' on the RHS, we obtain the (necessary) system

$$g'' + \frac{8\alpha^2}{\lambda^2} g = 0, \quad g(1) = 0, \;\; g'(0) = 0.$$

This is a second-order linear ODE. It has as its solutions

$$g(s) = A\cos\left(\frac{2\sqrt{2}\alpha}{|\lambda|}s\right) + B\sin\left(\frac{2\sqrt{2}\alpha}{|\lambda|}s\right)$$

for constants A, B to be determined with the boundary conditions. The condition $g'(0) = 0$ implies that $B = 0$. The condition $g'(1) = 0$ implies either that $A = 0$ or that

$$\frac{2\sqrt{2}\alpha}{|\lambda|} \in \frac{\pi}{2}(2\mathbb{Z}+1),$$

i.e., that the frequency is an odd multiple of $\pi/2$. This implies

$$|\lambda_k| = \frac{4\sqrt{2}\alpha}{\pi(2k+1)}, \quad k = 0, 1, \ldots,$$

and in particular

$$g_k(s) = A_k \cos\left(\frac{\pi}{2}(2k+1)s\right), \quad k = 0, 1, \ldots,$$

where the constants A_k can be determined such that g has unit L^2 norm. We have

$$\begin{aligned}\int_{-1}^{1} g_k(s) g_{k'}(s)\,\mathrm{d}s &= \frac{1}{2}\int_{-1}^{1} \left(\cos(\pi(k-k')s) + \cos(\pi(k+k'+1)s)\right)\mathrm{d}s \\ &= (\mathbb{1}_{k=k'} + \mathbb{1}_{k+k'+1=0}) \\ &= \mathbb{1}_{k=k'}.\end{aligned}$$

In particular, $A_k = 1$. It remains to determine the signs of the eigenvalues λ_k. We calculate

$$\begin{aligned}\bar{\mathcal{T}}_X[g_k](s) &= 2\sqrt{2}\alpha \int_0^{1-|s|} \cos\left(\frac{\pi}{2}(2k+1)s\right) \\ &= |\lambda_k| \sin\left(\frac{\pi}{2}(2k+1)(1-|s|)\right) \\ &= |\lambda_k| \sin\left(\frac{\pi}{2}(2k+1)\right) \cos\left(\frac{\pi}{2}(2k+1)s\right) \\ &= (-1)^k |\lambda_k| g_k(s).\end{aligned}$$

In particular, the functions g_k form a mutually orthogonal set of eigenfunctions of $\bar{\mathcal{T}}_X$ with corresponding eigenvalues

$$\lambda_k = (-1)^k \frac{4\sqrt{2}\alpha}{\pi(2k+1)}, \quad k = 0, 1, \ldots$$

To conclude, we note that from (F.19) that the functions $f_k : [-1, +1] \to \mathbb{R}$ defined by

$$f_k(s) = \begin{cases} \frac{1}{\sqrt{\alpha\sqrt{2}}} \cos\left(\frac{\pi}{2\sqrt{2}\alpha}(2k+1)s\right) & |s| \le \sqrt{2}\alpha \\ 0 & \text{otherwise,} \end{cases} \quad k = 0, 1, \ldots$$

form an orthonormal basis for the image of $\mathcal{T}_X$, and together with the eigenvalues λ_k defined above provide a Schmidt decomposition of the operator $\mathcal{T}_X$:

$$\mathcal{T}_X = \sum_{k \in \mathbb{N}_0} \lambda_k f_k f_k^*.$$

This completes the proof. □

Lemma F.2. *For all $m \in \mathbb{N}$, any $k \in [n]$, and any $\sigma^2 > 0$, one has for the operator defined in (F.11)*

$$\left\|\tilde{\boldsymbol{\Xi}}_m\right\|_{(k)}^{(2)} \le \frac{n}{2}\|\boldsymbol{\Lambda}_m\|_{(k)}^{(2)} + \frac{4m(1+\log m)^{1/2}}{\alpha} + \frac{\pi n \sigma^2 m^2}{32\sqrt{2}\alpha}.$$

Proof. We build from the matrix representation (F.11) of $\tilde{\boldsymbol{\Xi}}_m$. The idea of the proof is straightforward: if $\tilde{\boldsymbol{U}}_m$ had orthonormal columns, we would have immediately

$$\left\|\tilde{\boldsymbol{\Xi}}_m\right\|_{(k)}^{(2)} = \|\boldsymbol{\Lambda}_m\|_{(k)}^{(2)}, \tag{F.20}$$

by unitary invariance. Because of discretization and smoothing errors, $\tilde{\boldsymbol{U}}_m$ is not an orthonormal m-frame, so (F.20) does not hold. However, when n is large and m is not too large relative to n, we can guarantee that $\tilde{\boldsymbol{U}}_m$ is close to orthonormal, which we will combine with a perturbation result (Lemma F.16) to obtain the claim.

By Lemma F.16 and the triangle inequality, we have

$$\begin{aligned}\left\|\tilde{\boldsymbol{\Xi}}_m\right\|_{(k)}^{(2)} &\le \left\||\boldsymbol{\Lambda}_m|^{1/2}\tilde{\boldsymbol{U}}_m^*\tilde{\boldsymbol{U}}_m|\boldsymbol{\Lambda}_m|^{1/2}\right\|_{(k)}^{(2)} \\ &\le \tfrac{n}{2}\|\boldsymbol{\Lambda}_m\|_{(k)}^{(2)} + \left\||\boldsymbol{\Lambda}_m|^{1/2}\left(\tilde{\boldsymbol{U}}_m^*\tilde{\boldsymbol{U}}_m - \tfrac{n}{2}\boldsymbol{I}\right)|\boldsymbol{\Lambda}_m|^{1/2}\right\|_{(k)}^{(2)} \\ &\le \tfrac{n}{2}\|\boldsymbol{\Lambda}_m\|_{(k)}^{(2)} + \left\||\boldsymbol{\Lambda}_m|^{1/2}\left(\tilde{\boldsymbol{U}}_m^*\tilde{\boldsymbol{U}}_m - \tfrac{n}{2}\boldsymbol{I}\right)|\boldsymbol{\Lambda}_m|^{1/2}\right\|_{\mathrm{F}},\end{aligned} \tag{F.21}$$

where in the final inequality we worst-case the residual by summing over all singular values. We will bound the residual term in (F.21) by bounding the magnitude of each of its elements. For $j = 0, 1, \ldots, m-1$, let $\tilde{\boldsymbol{u}}_{m,j}$ denote the j-th column of $\tilde{\boldsymbol{U}}_m$, and let $\pi_1(G)$ denote the projection of the rectangle G onto its first coordinate. Then $(2/n)\langle \tilde{\boldsymbol{u}}_{m,j}, \tilde{\boldsymbol{u}}_{m,j'}\rangle$ is a Riemann sum corresponding to the integral of the function $(\varphi_{\sigma^2} * g_j)(\varphi_{\sigma^2} * g_{j'})$ over $[-1, 1]$. We have from the Leibniz rule

$$\begin{aligned}\|(\varphi_{\sigma^2} * g_j)(\varphi_{\sigma^2} * g_{j'})\|_{\mathrm{Lip}} &= \|\varphi_{\sigma^2} * g_j\|_{L^\infty}\|\varphi_{\sigma^2} * g_{j'}\|_{\mathrm{Lip}} + \|\varphi_{\sigma^2} * g_j\|_{\mathrm{Lip}}\|\varphi_{\sigma^2} * g_{j'}\|_{L^\infty} \\ &= \frac{1}{\sqrt{2}\alpha}\left(\|g_j\|_{\mathrm{Lip}} + \|g_{j'}\|_{\mathrm{Lip}}\right) \\ &= \frac{\pi(j+j'+1)}{2\alpha^2},\end{aligned}$$

where we use the fact that convolution with a gaussian does not increase L^∞ norms (a special case of Young's inequality [4, Ch. I, Thm 1.3]) nor Lipschitz seminorms (the functions g_j are all in C^∞, so we can differentiate under the integral and then apply Jensen's inequality, since the gaussian has unit L^1 norm). Thus, by Lemma F.15 and Lemma F.1 and the triangle inequality, we have

$$\begin{aligned}\left|\langle \tilde{\boldsymbol{u}}_{m,j}, \tilde{\boldsymbol{u}}_{m,j'}\rangle - \frac{n}{2}\mathbb{1}_{j=j'}\right| &\le \left|\langle \tilde{\boldsymbol{u}}_{m,j}, \tilde{\boldsymbol{u}}_{m,j'}\rangle - \frac{n}{2}\langle \varphi_{\sigma^2} * g_j, \varphi_{\sigma^2} * g_{j'}\rangle_{L^2}\right| + \left|\frac{n}{2}\langle \varphi_{\sigma^2} * g_j, \varphi_{\sigma^2} * g_{j'}\rangle_{L^2} - \frac{n}{2}\mathbb{1}_{j=j'}\right| \\ &\le \frac{n}{2}|\langle \varphi_{\sigma^2} * g_j, \varphi_{\sigma^2} * g_{j'}\rangle_{L^2} - \mathbb{1}_{j=j'}| + \frac{\pi(j+j'+1)}{2\alpha^2}. \end{aligned} \tag{F.22}$$

To handle the remaining residual, we will apply Lemma F.18. This gives

$$\frac{n}{2}|\langle \varphi_{\sigma^2} * g_j, \varphi_{\sigma^2} * g_{j'}\rangle_{L^2} - \mathbb{1}_{j=j'}| \le \frac{n\sigma^2}{2}\|g_j'\|_{L^2}\|g_{j'}'\|_{L^2},$$

and from Lemma F.1 and a L^1-L^∞ estimate, we have

$$\|g_j'\|_{L^2} \le \frac{\pi(2j+1)}{2\sqrt{2}\alpha},$$

whence

$$\frac{n}{2}|\langle \varphi_{\sigma^2} * g_j, \varphi_{\sigma^2} * g_{j'}\rangle_{L^2} - \mathbb{1}_{j=j'}| \le \frac{n\sigma^2\pi^2(2j+1)(2j'+1)}{16\alpha^2}.$$

In particular, substituting this estimate into (F.22) gives

$$\left|\langle \tilde{\boldsymbol{u}}_{m,j}, \tilde{\boldsymbol{u}}_{m,j'}\rangle - \frac{n}{2}\mathbb{1}_{j=j'}\right| \le \frac{n\sigma^2\pi^2(2j+1)(2j'+1)}{16\alpha^2} + \frac{\pi(j+j'+1)}{2\alpha^2}. \tag{F.23}$$

From the definition of $\boldsymbol{\Lambda}_m$, it follows

$$\left\||\boldsymbol{\Lambda}_m|^{1/2}\left(\boldsymbol{U}_m^*\boldsymbol{U}_m - \tfrac{n}{2}\boldsymbol{I}\right)|\boldsymbol{\Lambda}_m|^{1/2}\right\|_{\mathrm{F}}^2 \le \frac{16}{\alpha^2}\sum_{1\le i,j\le m}\frac{(i+j-1)^2}{(2i-1)(2j-1)} + \frac{\pi^2n^2\sigma^4}{2^{11}\alpha^2}\sum_{1\le i,j\le m}(2i-1)(2j-1).$$

The second sum evaluates to m^4. For the first sum, note that $i+j-1 = \frac{1}{2}((2i-1)+(2j-1))$, so

$$\begin{aligned}\frac{(i+j-1)^2}{(2i-1)(2j-1)} &= \frac{1}{4}\left(\sqrt{\frac{2i-1}{2j-1}} + \sqrt{\frac{2j-1}{2i-1}}\right)^2 \\ &\le \frac{1}{2}\left(\frac{2i-1}{2j-1} + \frac{2j-1}{2i-1}\right),\end{aligned}$$

by the inequality $a+b \le 2(a^2+b^2)$. When summed over the grid $[m]^2$, the two functions of i, j in the last inequality must be equal by symmetry. Thus

$$\begin{aligned}\sum_{1\le i,j\le m}\frac{(i+j-1)^2}{(2i-1)(2j-1)} &\le \sum_{1\le i,j\le m}\frac{2i-1}{2j-1} \\ &= m^2\sum_{j=1}^m\frac{1}{2j-1}.\end{aligned}$$

By the usual estimates $\log m \le \sum_{j=1}^m \frac{1}{j} \le 1+\log m$ for the harmonic numbers, we have $\sum_{j=1}^m \frac{1}{2j-1} \le 1+\log(2m)-\frac{1}{2}\log m$, which in turn is less than $1+\log m$ when $m \ge 4$. In addition, one can check numerically that the same estimate holds for $m \in \{1,2,3,4\}$. We have thus shown

$$\left\||\boldsymbol{\Lambda}_m|^{1/2}\left(\boldsymbol{U}_m^*\boldsymbol{U}_m - \tfrac{n}{2}\boldsymbol{I}\right)|\boldsymbol{\Lambda}_m|^{1/2}\right\|_{\mathrm{F}}^2 \le \frac{8m^2(1+\log m)}{\alpha^2} + \frac{\pi^2n^2\sigma^4m^4}{2^{11}\alpha^2},$$

which establishes the claim when combined with our previous estimates.

□

Lemma F.3. *For all $m \in \mathbb{N}$ and any $k \in [n]$, one has for the operator defined in (F.11)*

$$\|\mathbf{\Lambda}_m\|_{(k)}^{(2)} \leq \sqrt{4\alpha^2 - \frac{16\alpha^2}{\pi^2}\frac{1}{2\min\{m,k\}+1}}.$$

Proof. We have by definition

$$\begin{aligned}\left(\|\mathbf{\Lambda}_m\|_{(k)}^{(2)}\right)^2 &= \frac{32\alpha^2}{\pi^2}\sum_{i=1}^{\min\{k,m\}}\frac{1}{(2k-1)^2} \\ &= \frac{32\alpha^2}{\pi^2}\left(\sum_{i=1}^{\infty}\frac{1}{(2k-1)^2} - \sum_{i=1+\min\{k,m\}}^{\infty}\frac{1}{(2k-1)^2}\right). \end{aligned} \tag{F.24}$$

For the first term, we have

$$\begin{aligned}\sum_{i=1}^{\infty}\frac{1}{(2k-1)^2} &= \sum_{i=1}^{\infty}\frac{1}{k^2} - \sum_{i=1}^{\infty}\frac{1}{(2k)^2} \\ &= \frac{3}{4}\sum_{i=1}^{\infty}\frac{1}{k^2} \\ &= \frac{\pi^2}{8}.\end{aligned}$$

For the second term, we have from the integral test estimate

$$\begin{aligned}\sum_{i=1+\min\{k,m\}}^{\infty}\frac{1}{(2k-1)^2} &\geq \int_{1+\min\{k,m\}}^{\infty}\frac{1}{(2t-1)^2}\,\mathrm{d}t \\ &= \frac{\frac{1}{2}}{2\min\{k,m\}+1}.\end{aligned}$$

Plugging into (F.24) and taking square roots gives the claim. □

Lemma F.4. *Consider the smoothed template $\bar{X}$, as in Theorem F.1, with sampling $\bar{\boldsymbol{X}}$ on the grid G. Let $\boldsymbol{X}$ denote the "directly sampled" template*

$$(\boldsymbol{X})_{ij} = \mathbb{1}_{\|(\boldsymbol{\tau}_{\pi/4})_{ij}\|_\infty \leq \alpha},$$

where we recall (F.96) and (F.97). Then if $\sigma \leq \frac{1}{2}$, one has

$$\|\bar{\boldsymbol{X}} - \boldsymbol{X}\|_{\ell^2(G)}^2 \leq \frac{n^2\sigma^8}{\pi^2} + \frac{2n}{\pi} + \frac{n^2\sqrt{48\sigma^2\log(1/\sigma)}}{\pi}.$$

Proof. Note that by definition

$$\bar{X}(i,j) = \int_{\mathbb{R}^2}\varphi_{\sigma^2}((i,j)-\boldsymbol{x}')X(\boldsymbol{x}')\,\mathrm{d}\boldsymbol{x}'.$$

Because

$$\varphi_{\sigma^2}(\boldsymbol{x}) = \frac{1}{2\pi\sigma^2}e^{-\frac{1}{2\sigma^2}\|\boldsymbol{x}\|_2^2},$$

for $\|\boldsymbol{x}\|_2^2 \geq 12\sigma^2\log(1/\sigma)$, one has

$$\varphi_{\sigma^2}(\boldsymbol{x}) \leq \frac{\sigma^4}{2\pi}. \tag{F.25}$$

Since $\|\boldsymbol{x}\|_2 \geq \|\boldsymbol{x}\|_\infty$ and $\|\boldsymbol{R}_\nu\boldsymbol{x}\|_2 = \|\boldsymbol{x}\|_2$, if $\|\boldsymbol{R}_{\pi/4}\boldsymbol{x}\|_\infty^2 \geq 12\sigma^2\log(1/\sigma)$, then (F.25) also holds.

Consider the set

$$S = \left\{(i,j) \in G \;\middle|\; \left|\left\|(\boldsymbol{\tau}_{\pi/4})_{ij}\right\|_\infty - \alpha\right| > \sqrt{12\sigma^2\log(1/\sigma)}\right\}, \tag{F.26}$$

and write S^{c} for the complement of S relative to G. First, suppose that $(i,j) \in S$ is not in the support of X. We have $X(i,j) = 0$, and

$$\begin{aligned}
\bar{X}(i,j) &= \int_{\mathbb{R}^2} \varphi_{\sigma^2}((i,j) - \boldsymbol{x}')X(\boldsymbol{x}')\,\mathrm{d}\boldsymbol{x}' \\
&= \int_{\|\boldsymbol{R}_{\pi/4}\boldsymbol{x}\|_\infty \le \alpha} \varphi_{\sigma^2}((i,j) - \boldsymbol{x}')X(\boldsymbol{x}')\,\mathrm{d}\boldsymbol{x}' + \int_{\|\boldsymbol{R}_{\pi/4}\boldsymbol{x}\|_\infty \ge \alpha} \varphi_{\sigma^2}((i,j) - \boldsymbol{x}')X(\boldsymbol{x}')\,\mathrm{d}\boldsymbol{x}' \\
&\le \frac{\sigma^4}{2\pi} \int_{\|\boldsymbol{R}_{\pi/4}\boldsymbol{x}\|_\infty \le \alpha} X(\boldsymbol{x}')\,\mathrm{d}\boldsymbol{x}' \\
&= \frac{\sigma^4}{\pi}
\end{aligned}$$

by the triangle inequality. Evidently $\bar{X}(i,j) \ge 0$ as well. A symmetric argument applies when $(i,j) \in S$ is in the support of X, except that we obtain

$$\begin{aligned}
\bar{X}(i,j) &= \int_{\mathbb{R}^2} \varphi_{\sigma^2}(\boldsymbol{x}')X((i,j) - \boldsymbol{x}')\,\mathrm{d}\boldsymbol{x}' \\
&= \int_{\|\boldsymbol{R}_{\pi/4}((i,j)-\boldsymbol{x}')\|_\infty \le \alpha} \varphi_{\sigma^2}(\boldsymbol{x}')\,\mathrm{d}\boldsymbol{x}' \\
&= 1 - \int_{\|\boldsymbol{R}_{\pi/4}((i,j)-\boldsymbol{x}')\|_\infty \ge \alpha} \varphi_{\sigma^2}(\boldsymbol{x}')\,\mathrm{d}\boldsymbol{x}',
\end{aligned}$$

whence

$$\begin{aligned}
\left|\bar{X}(i,j) - 1\right| &\le \int_{\|\boldsymbol{R}_{\pi/4}((i,j)-\boldsymbol{x}')\|_\infty \ge \alpha} \varphi_{\sigma^2}(\boldsymbol{x}')\,\mathrm{d}\boldsymbol{x}' \\
&\le \int_{\|\boldsymbol{R}_{\pi/4}\boldsymbol{x}'\|_\infty \ge \sqrt{12\sigma^2\log(1/\sigma)}} \varphi_{\sigma^2}(\boldsymbol{x}')\,\mathrm{d}\boldsymbol{x}' \\
&\le \int_{\|\boldsymbol{x}'\|_2 \ge \sqrt{6\sigma^2\log(1/\sigma)}} \varphi_{\sigma^2}(\boldsymbol{x}')\,\mathrm{d}\boldsymbol{x}'.
\end{aligned}$$

The last integral can be estimated with the fact that the gaussian tail integral is bounded by the density. This gives

$$\left|\bar{X}(i,j) - 1\right| \le \frac{\sigma^6}{24\pi\log(1/\sigma)} \le \frac{\sigma^4}{\pi},$$

where the last simplification holds if $\sigma \le \frac{1}{2}$. Thus, we have shown that for any $(i,j) \in S$, we have

$$|X(i,j) - \bar{X}(i,j)| \le \sigma^4/\pi. \tag{F.27}$$

Next, we argue that the cardinality $|S|$ is sufficiently large. We will do this by bounding the size of S^{c}. We have by inequalities for ℓ^p norms and (F.96) and (F.97)

$$\frac{1}{\sqrt{2}}\left\|\begin{bmatrix} i \\ j \end{bmatrix}\right\|_2 \le \|\boldsymbol{\tau}_{\nu ij}\|_\infty \le \left\|\begin{bmatrix} i \\ j \end{bmatrix}\right\|_2,$$

so if we define

$$S' = \left\{(i,j) \in G \;\middle|\; \left|\left\|\begin{bmatrix} i \\ j \end{bmatrix}\right\|_2 - \alpha\right| \le \sqrt{24\sigma^2\log(1/\sigma)}\right\},$$

we have $S^{\mathsf{c}} \subset S'$. The square $[-1,1]$ is covered by the union of balls of radius $\sqrt{2}/(n-1)$ centered at each point of the grid G. Consider the subset

$$U = \left\{(u,v) \in [-1,1] \;\middle|\; \left|\sqrt{u^2+v^2} - \alpha\right| \le \frac{\sqrt{2}}{n-1} + \sqrt{24\sigma^2\log(1/\sigma)}\right\}.$$

Then by the triangle inequality and the above covering reasoning, $S' + \{(u,v) \in \mathbb{R}^2 \mid \sqrt{u^2+v^2} \le \sqrt{2}/(n-1)\} \subset U$, from which it follows by a volume bound

$$|S^{\mathsf{c}}|\pi\left(\frac{\sqrt{2}}{n-1}\right)^2 \le \mathrm{Vol}(U).$$

We calculate readily

$$\mathrm{Vol}(U) = 4\alpha\left(\frac{\sqrt{2}}{n-1} + \sqrt{24\sigma^2\log(1/\sigma)}\right),$$

whence

$$|S^{\mathsf{c}}| \le \frac{2n}{\pi} + \frac{n^2\sqrt{48\sigma^2\log(1/\sigma)}}{\pi},$$

where the last inequality worst-cases over our condition on α.

Now, to conclude, we have by the above

$$\begin{aligned}\left\|\bar{\boldsymbol{X}} - \boldsymbol{X}\right\|_{\mathrm{F}}^2 &= \sum_{(i,j)\in S}\left((\bar{\boldsymbol{X}})_{ij} - (\boldsymbol{X}\circ\boldsymbol{\tau}_\nu)_{ij}\right)^2 \sum_{(i,j)\in S^{\mathsf{c}}}\left((\bar{\boldsymbol{X}})_{ij} - (\boldsymbol{X}\circ\boldsymbol{\tau}_\nu)_{ij}\right)^2 \\ &\le \frac{n^2\sigma^8}{\pi^2} + |S^{\mathsf{c}}|\sup_{(i,j)\in G}\left((\bar{\boldsymbol{X}})_{ij} - (\boldsymbol{X}\circ\boldsymbol{\tau}_\nu)_{ij}\right)^2 \\ &\le \frac{n^2\sigma^8}{\pi^2} + \frac{2n}{\pi} + \frac{n^2\sqrt{48\sigma^2\log(1/\sigma)}}{\pi},\end{aligned}$$

because both matrices have entries in $[0,1]$.

□

Lemma F.5. *For $\nu = \pi/4$, consider the "directly sampled" infinite-resolution template*

$$(\bar{\boldsymbol{X}})_{ij} = \mathbb{1}_{\|(\boldsymbol{\tau}_\nu)_{ij}\|_\infty \le \alpha}, \tag{F.28}$$

where we recall (F.96) and (F.97). Then one has

$$\left\|\bar{\boldsymbol{X}}\right\|_{\mathrm{F}} \ge n^2\alpha^2 - 5n.$$

Proof. Consider the case $\nu = \pi/4$. By rotational symmetry of $\bar{\boldsymbol{X}}$ by multiples of $\pi/2$, and discarding the sum over the central axes when n is odd, we have

$$\begin{aligned}\left\|\bar{\boldsymbol{X}}\right\|_{\mathrm{F}}^2 &= \sum_{(i,j)\in G}\mathbb{1}_{\max\{|i+j|,|i-j|\}\le\sqrt{2}\alpha} \\ &= \sum_{i=0}^{n-1}\sum_{j=0}^{n-1}\mathbb{1}_{\max\{|i-(n-1-j)|,|i-j|\}\le(n-1)\alpha/\sqrt{2}} \\ &\ge 4\sum_{i=0}^{\lfloor\frac{n-1}{2}\rfloor}\sum_{j=0}^{\lfloor\frac{n-1}{2}\rfloor}\mathbb{1}_{|i-(n-1-j)|\le(n-1)\alpha/\sqrt{2}}.\end{aligned}$$

So, by the integral test estimate (because the summand is monotone increasing as a function of both i and j when the other is

fixed) and nonnegativity,

$$\begin{aligned}
\left\|\bar{\boldsymbol{X}}\right\|_{\mathrm{F}}^2 &\geq 4\int_0^{\lfloor\frac{n-1}{2}\rfloor}\int_0^{\lfloor\frac{n-1}{2}\rfloor} \mathbb{1}_{|i-(n-1-j)|\leq(n-1)\alpha/\sqrt{2}}\,\mathrm{d}i\,\mathrm{d}j \\
&= 4(n-1)^2\int_0^{\frac{1}{n-1}\lfloor\frac{n-1}{2}\rfloor}\int_0^{\frac{1}{n-1}\lfloor\frac{n-1}{2}\rfloor} \mathbb{1}_{|i-(1-j)|\leq\alpha/\sqrt{2}}\,\mathrm{d}i\,\mathrm{d}j \\
&\geq 4(n-1)^2\int_0^{\frac{1}{2}-\frac{1}{n-1}}\int_0^{\frac{1}{2}-\frac{1}{n-1}} \mathbb{1}_{|i-(1-j)|\leq\alpha/\sqrt{2}}\,\mathrm{d}i\,\mathrm{d}j \\
&\geq 4(n-1)^2\left(\int_0^{\frac{1}{2}}\int_0^{\frac{1}{2}} \mathbb{1}_{|i-(1-j)|\leq\alpha/\sqrt{2}}\,\mathrm{d}i\,\mathrm{d}j - 2\int_0^{\frac{1}{2}}\int_{\frac{1}{2}-\frac{1}{n-1}}^{\frac{1}{2}} \mathbb{1}_{|i-(1-j)|\leq\alpha/\sqrt{2}}\,\mathrm{d}i\,\mathrm{d}j\right),
\end{aligned}$$

where in the final inequality we used permutation symmetry of the integral as a function of (i,j) to simplify the residual. Now, the region of integration in the first term in the last line of the previous expression is equivalent to $\{(i,j)\,|\,(\frac{1}{2}-i)+(\frac{1}{2}-j)\leq \alpha/\sqrt{2}\}$, which defines a right triangle with two side lengths equal to $\alpha/\sqrt{2}$. Because $\alpha < 1/\sqrt{2}$, the integral evaluates to $\alpha^2/4$. Meanwhile, the integral in the second term is no larger than $1/2(n-1)$, because the integrand is bounded by 1. Thus

$$\left\|\bar{\boldsymbol{X}}\right\|_{\mathrm{F}}^2 \geq (n-1)^2\alpha^2 - 4(n-1).$$

Distributing in this expression and worst-casing slightly gives the claim.

□

F.2. Proofs for Theorem 2

Problem setup (and continuum idealization). Let $k\in\mathbb{N}$, and consider an observation $\boldsymbol{X}\in\mathbb{R}^{m\times n}$ drawn from the class (F.3), with rotation parameter $\nu_\natural$ (so that $(\boldsymbol{X})_{ij} = X_\natural\circ\boldsymbol{\tau}_{-\nu_\natural}(i,j)$ if $(i,j)\in G$, following (F.6)). For $\boldsymbol{U}\in\mathbb{R}^{m\times k}$, $\boldsymbol{V}\in\mathbb{R}^{n\times k}$, we study the optimization objective

$$\mathcal{L}_{\text{discrete}}(\nu,\boldsymbol{U},\boldsymbol{V}) = \frac{1}{2}\|\boldsymbol{X} - (\boldsymbol{U}\boldsymbol{V}^*)\circ\boldsymbol{\tau}_{-\nu}\|_{\mathrm{F}}^2, \tag{F.29}$$

where here in the context of discrete images, the transformations $\boldsymbol{\tau}_{-\nu}$ must be implemented with resampling (we give a brief overview of this idea in Section F.4). The resampling operation can be chosen to be continuously differentiable, making it amenable to gradient-based optimization on the objective $\mathcal{L}_{\text{discrete}}$, but it introduces a host of discretization-based artifacts to the optimization process that are challenging to treat.[2] We will simplify the situation by considering a continuum limit of the objective (F.29), and a corresponding continuum gradient-like iteration for its solution.

Consider operators $\boldsymbol{U}:\mathbb{R}^k\to L^2(\mathbb{R})$, $\boldsymbol{V}:\mathbb{R}^k\to L^2(\mathbb{R})$. These operators can be thought of as 'matrices', whose columns are $L^2(\mathbb{R})$ functions—note that in the continuum, following (F.5) we have

$$X_\natural(s,t) = \underbrace{\mathbb{1}_{|s|\leq\alpha}}_{u_\natural(s)}\underbrace{\mathbb{1}_{|t|\leq\alpha}}_{v_\natural(t)},$$

i.e., as an operator, $X_\natural = u_\natural v_\natural^*$. The corresponding continuum analogue of the observation $\boldsymbol{X}$ is the deformed template $X = X_\natural\circ\boldsymbol{\tau}_{-\nu_\natural}$. To mirror the smoothing effect of a continuous interpolation kernel imposed in (F.95), we introduce an extra gaussian smoothing filter $\varphi_{\sigma^2}(s,t) = (2\pi\sigma^2)^{-1}e^{-(s^2+t^2)/2\sigma^2}$ to the objective function, yielding the objective

$$\mathcal{L}^\sigma(\nu,\boldsymbol{U},\boldsymbol{V}) = \frac{1}{2}\|\varphi_{\sigma^2} * (X - (\boldsymbol{U}\boldsymbol{V}^*)\circ\boldsymbol{\tau}_{-\nu})\|_{L^2}^2. \tag{F.30}$$

[2] In particular, note that for any ν, $\boldsymbol{M}\mapsto\boldsymbol{M}\circ\boldsymbol{\tau}_\nu$, as defined in Section F.4, is a linear operator. If we call this operator $\mathcal{A}_\nu$, it can be seen from Section F.4 that $\mathcal{A}_\nu^*\mathcal{A}_\nu$ is a banded matrix, but it is *not* incoherent—this means that the analysis of the problem (F.29) requires tools other those developed to analyze matrix sensing under the RIP (c.f. [18, 29]). The situation is further complicated by the fact that the objective (F.29) simultaneously learns the sensing matrix (in matrix sensing parlance) and the low-rank factorization, a setting that has not been considered in prior work.

Notice that, by its definition (F.2), the map $f \mapsto f \circ \boldsymbol{\tau}_{-\nu}$ is a unitary transformation (apply the change of variables formula in the integral defining the L^2 inner product). Using in addition the Lie group structure of the rotation matrices (F.2), we have that for any $f \in L^2(\mathbb{R}^2)$, any ν and any σ^2,

$$\begin{aligned}(\varphi_{\sigma^2} * (f \circ \boldsymbol{\tau}_\nu))(\boldsymbol{x}) &= \int_{\mathbb{R}^2} \varphi_{\sigma^2}(\boldsymbol{x}') f(\boldsymbol{R}_\nu(\boldsymbol{x} - \boldsymbol{x}'))\,\mathrm{d}\boldsymbol{x}' \\ &= \int_{\mathbb{R}^2} \varphi_{\sigma^2}(\boldsymbol{R}_{-\nu}(\boldsymbol{x}')) f(\boldsymbol{R}_\nu(\boldsymbol{x}) - \boldsymbol{x}')\,\mathrm{d}\boldsymbol{x}' \\ &= \int_{\mathbb{R}^2} \varphi_{\sigma^2}(\boldsymbol{x}') f(\boldsymbol{R}_\nu(\boldsymbol{x}) - \boldsymbol{x}')\,\mathrm{d}\boldsymbol{x}' \\ &= (\varphi_{\sigma^2} * f) \circ \boldsymbol{\tau}_\nu(\boldsymbol{x}) \end{aligned} \tag{F.31}$$

by the change of variables formula and rotational invariance of the gaussian function. In words, rigid motions commute with gaussian smoothing. Applying this result together with the unitary transformation property, we can write our objective as

$$\begin{aligned}\mathcal{L}^\sigma(\nu, \boldsymbol{U}, \boldsymbol{V}) &= \frac{1}{2}\|\varphi_{\sigma^2} * X - (\varphi_{\sigma^2} * (\boldsymbol{U}\boldsymbol{V}^*)) \circ \boldsymbol{\tau}_{-\nu}\|_{L^2}^2 \\ &= \frac{1}{2}\|(\varphi_{\sigma^2} * X) \circ \boldsymbol{\tau}_\nu - \varphi_{\sigma^2} * (\boldsymbol{U}\boldsymbol{V}^*)\|_{L^2}^2 \\ &= \frac{1}{2}\left\|\varphi_{\sigma^2} * \left(X_\natural \circ \boldsymbol{\tau}_{\nu - \nu_\natural} - \boldsymbol{U}\boldsymbol{V}^*\right)\right\|_{L^2}^2. \end{aligned} \tag{F.32}$$

We emphasize that (F.30) and (F.32) are equal, but (F.32) is more straightforward to differentiate.

Simplifications to (F.32). Our analysis will apply to a simplified version of the general objective (F.32). We discuss the simplifications we make here.

1. **Single-channel factorization ($k = 1$).** We analyze a critically-parameterized version of the problem (F.32), where $k = 1$. This leads to the objective

$$\mathcal{L}^\sigma(\nu, u, v) = \frac{1}{2}\left\|\varphi_{\sigma^2} * \left(X_\natural \circ \boldsymbol{\tau}_{\nu - \nu_\natural} - uv^*\right)\right\|_{L^2}^2, \tag{F.33}$$

   where $u, v \in L^2(\mathbb{R})$. When the transformation component of (F.33) is omitted, this simplification is analogous to consideration of the rank-one matrix factorization problem [18, §3]; because the untransformed square template $X_\natural$ has "rank one" (in a suitable, generalized sense), perfect reconstruction is still possible in our setting. Although the rank-one case is a vast simplification over the problem (F.32) with general k, we begin our analysis here because the introduction of the simultaneous transformation optimization component to (F.33) represents a nontrivial complication with respect to existing analyses (c.f. [15, 22]). We anticipate that the emerging understanding of overparameterized matrix sensing will be useful in generalizing our results to the setting of general k [16, 29, 37, 38].

2. **Symmetric factorization.** Because the square template $X_\natural$ is self-adjoint as an integral operator (in other words, the template satisfies $X_\natural(s,t) = X_\natural(t,s)$), it is reasonable to reduce the search space in (F.33) to factorizations where $u = v$. This gives the problem

$$\mathcal{L}^\sigma(\nu, u) = \frac{1}{2}\left\|\varphi_{\sigma^2} * \left(X_\natural \circ \boldsymbol{\tau}_{\nu - \nu_\natural} - uu^*\right)\right\|_{L^2}^2. \tag{F.34}$$

   All of our experiments with TILTED make use of general, asymmetric grid factors, so a theoretical understanding of the general asymmetric case (when the target template $X_\natural$ is asymmetric) remains crucial for future work. In this connection, we note that theoretical analyses of asymmetric matrix factorization typically add an additional "balancing" regularizer to the objective (F.32) (c.f. [14, 15, 18, 22])—in our experiments, the 2, 1 regularizer described in Section A.3 plays this role.

Gradient-like iterations for alignment and factorization. Obtaining a gradient iteration for the objective (F.32) can be done straightforwardly with respect to the finite-dimensional ν variable: making essential use of the duality on $L^2(\mathbb{R})$ and the

fact that the convolution of two gaussian functions is another gaussian function, with variance equal to the sum of the factors' variances, we calculate in Lemma F.7

$$\nabla_\nu \mathcal{L}^\sigma(\nu, u) = -\left\langle \varphi_{\sigma^2} * \left(uu^* \circ \boldsymbol{\tau}_{\nu_\natural - \nu}\right), \left\langle \nabla_{\boldsymbol{x}}[\varphi_{\sigma^2} * X_\natural], \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} (\,\cdot\,) \right\rangle_{\ell^2} \right\rangle_{L^2(\mathbb{R}^2)}. \tag{F.35}$$

Differentiation with respect to the u factor in (F.34) requires a slightly more technical notion of gradient. To limit technicality in the analysis, we study instead an infinite-dimensional analogue of a projected gradient descent method, where after each update to the u variable we project it onto the "unit sphere" in $L^2(\mathbb{R})$ as $u \mapsto u/\|u\|_{L^2(\mathbb{R})}$. Moreover, when performing factorization, we optimize only the unsmoothed loss $\mathcal{L}^0(\nu, \,\cdot\,)$. We recall in Lemma F.19 that in this setting, whenever the factorization target $X_\natural \circ \boldsymbol{\tau}_{\nu - \nu_\natural}$ is not negative, it is equivalent in this setting to seek the largest positive eigenvalue of the (symmetrized) operator corresponding to the factorization target. Moreover, as long as the factorization target has no negative eigenvalues of significant magnitude,[3] this process is achieved by the power method, which in our setting produces iterates

$$u_{k+1} = \frac{\left(\mathcal{T}_{X_\natural \circ \boldsymbol{\tau}_{\nu - \nu_\natural}} + \mathcal{T}^*_{X_\natural \circ \boldsymbol{\tau}_{\nu - \nu_\natural}}\right) u_k}{\left\| \left(\mathcal{T}_{X_\natural \circ \boldsymbol{\tau}_{\nu - \nu_\natural}} + \mathcal{T}^*_{X_\natural \circ \boldsymbol{\tau}_{\nu - \nu_\natural}}\right) u_k \right\|_{L^2(\mathbb{R})}}, \tag{F.36}$$

before outputting an approximate factor for the target, which we will write as $\mathsf{P}(k, u_0, \nu) \in L^2(\mathbb{R})$ (specifying the dependence on the power method's initialization u_0 and the rotation ν applied to the template):

$$\mathsf{P}(k, u_0, \nu) = \sqrt{\tfrac{1}{2} u_k^* (\mathcal{T}_{X_\natural \circ \boldsymbol{\tau}_{\nu - \nu_\natural}} + \mathcal{T}^*_{X_\natural \circ \boldsymbol{\tau}_{\nu - \nu_\natural}}) u_k}\, u_k. \tag{F.37}$$

A priori, this may return a complex-valued function; we will take care in our analysis to show that this never occurs when the iteration count is set appropriately.

Our algorithm. The key mathematical property underlying the success of TILTED in practical experiments is the fact that *incremental improvements to representation (factorization) help promote incremental improvements to alignment, and vice versa*. The algorithm we study theoretically is a simplified version of TILTED, but nonetheless captures this complex interplay and sheds light on why it succeeds in practice. The major simplification we impose is that rather than jointly updating the ν (alignment) iterates and the u (factorization) iterates, we will update them individually in consecutive blocks, as in an alternating minimization procedure. Our algorithm separates into five distinct stages, described below.

Stage one: rough representation. From a "flat" initialization for the scene

$$u_0 = \mathbb{1}_{[-1,+1]}, \tag{F.38}$$

we perform T_{rough} iterations of power method (F.37), to generate a roughly-localized representation of the template $X_\natural$:

$$u_{\text{rough}} = \mathsf{P}(T_{\text{rough}}, u_0, 0). \tag{F.39}$$

This procedure corresponds to the initial iterations of TILTED in practical experiments, where the uninformative initialization u_0 does not produce sufficient gradients (in texture or geometry) for alignment to occur. The roughly-localized output u_{rough} usefully ends up with both *texture* and a rough *shape* profile that promotes subsequent alignment.

Stage two: rough alignment. Given a step size β, we perform T_ν iterations of gradient descent on the alignment objective with smoothing level σ, initialized randomly:

$$\nu_0 \sim \text{Unif}([0, 2\pi]), \tag{F.40}$$

and with the factorization iterate at the output of the previous rough representation step:

$$\nu_{k+1} = \nu_k - \beta \nabla_\nu \mathcal{L}^\sigma(\nu_k, u_{\text{rough}}), \qquad k = 0, 1, \ldots, T_\nu - 1. \tag{F.41}$$

[3] This is typically the case; see Lemma F.1. Our proofs show that this structure persists throughout iterations of our algorithm, as well, in order to guarantee that our algorithm succeeds.

We write $\hat{\nu} = \nu_{T_\nu}$. The alignment problem in (F.34) has multiple optimal solutions, due to the symmetries of the alignment target $X_\natural$—this means the optimization landscape is *not* globally convex. At these initial iterations of the alignment procedure, with a non-informative initialization, we rely on the presence of strong gradient (mnemonically, "SG") in the objective landscape to bring our initial iterate close to one of the several equivalent optimal solutions. At a technical level, this style of analysis mirrors those used in other global analyses of nonconvex optimization landscapes [22].

Stage three: refined representation. This final stage of the algorithm takes advantage of the roughly-localized alignment output from the previous stage to improve the representation quality further—the initial roughly-localized template u_{rough} from the first stage is better localized, and its edges sharpened to match those of the target $X_\natural$. Accordingly, we run T_u iterations of power method (F.37), started with the outputs of the previous stages:

$$\hat{u} = \mathsf{P}(T_u, u_{\text{rough}}, \hat{\nu}). \tag{F.42}$$

The algorithm's output is the pair $(\hat{\nu}, \hat{u})$.

Main result. Our main result establishes convergence of our alternating minimization version of TILTED to the true parameters $(\nu_\natural, u_\natural)$, up to symmetry, in a "hard" instance of the problem: where $\nu_\natural = \pi/4$ (as we studied in Section F.1) and $\alpha = \frac{1}{\sqrt{2}}$ (corresponding to an 'in focus' target).

Theorem F.2. *Consider the iterations encompassed by* (F.39), (F.42) *and* (F.41), *with initializations* (F.38) *and* (F.40). *Suppose* $\alpha = \frac{1}{\sqrt{2}}$ *and* $\nu_\natural = \pi/4$. *There are absolute constants* c_1, C_1, $C_2 > 0$ *such that for any parameters* σ, β *satisfying*

$$\begin{aligned} \sigma^2 &\le \tfrac{1}{10^4}, \\ \beta &\le c_1, \end{aligned}$$

for any $0 < \varepsilon \le \frac{1}{768}$, *if the iteration counts satisfy*

$$\begin{aligned} T_{\text{rough}} &\ge -C_1 \log(\sigma^2 \varepsilon), \\ T_\nu &\ge -\frac{C_2 \log(3\varepsilon)}{\beta}, \\ T_u &\ge 16, \end{aligned}$$

then with probability over the random initialization of ν_0 *at least* $4/7$, *one has*

$$\begin{aligned} \min\left\{ |\hat{\nu} - \nu_\natural| \mod \frac{\pi}{2}, \frac{\pi}{2} - \left(|\hat{\nu} - \nu_\natural| \mod \frac{\pi}{2} \right) \right\} &\le 3\varepsilon, \\ \|\hat{u} - u_\natural\|_{L^2(\mathbb{R})} &\le 31\sqrt{\varepsilon}. \end{aligned}$$

In particular, the template parameters are recovered up to symmetry.

Proof. Following the alternating structure of the algorithm we study, the proof separates into a distinct stage for each phase of the algorithm. For concision, we will not carefully track the value of absolute constants in some parts of the proof; expressions such as $c, c_1, \dots$ and $C, C_1, \dots$ will denote small (respectively, large) absolute constants whose value may change from line to line unless otherwise noted. We will also use the expression $f \lesssim g$ to denote the statement "there exists an absolute constant $C > 0$ such that $f \le Cg$" for functions f, g, and analogously for $f \gtrsim g$.

Rough factorization stage. We will apply Lemma F.6 to the final iterate (F.39), obtained via the power method (F.37); to this end, we need to check properties of the operator

$$\frac{1}{2}\left(\mathcal{T}_{X_\natural \circ \boldsymbol{\tau}_{\pi/4}} + \mathcal{T}_{X_\natural \circ \boldsymbol{\tau}_{\pi/4}} \right) = \mathcal{T}_{X_\natural \circ \boldsymbol{\tau}_{\pi/4}}$$

(the simplification uses symmetry properties of the $\pi/4$-rotated template), and of the initialization (F.38). Notice that by Lemma F.1, we have

$$\left\| \mathcal{T}_{X_\natural \circ \boldsymbol{\tau}_{\pi/4}} \right\| = \lambda_{\max}\left(\mathcal{T}_{X_\natural \circ \boldsymbol{\tau}_{\pi/4}} \right) = \frac{4}{\pi},$$

which we will denote as λ_1 (in the notation of Lemma F.6), and its corresponding unit eigenvector is $v_1(s) = \cos(\pi s/2)\mathbb{1}_{|s|\le 1}$. Moreover, Lemma F.1 shows that the sequence of eigenvalue magnitudes of this operator are a decreasing function of index k, and therefore

$$\left|\frac{\lambda_k}{\lambda_1}\right| \le \frac{4}{3\pi}\cdot\frac{\pi}{4} = 1 - \frac{2}{3}.$$

In addition, we calculate

$$\langle u_0, v_1\rangle_{L^2(\mathbb{R})} = \langle \mathbb{1}_{[-1,1]}, \cos(\pi s/2)\rangle_{L^2(\mathbb{R})} = \frac{4}{\pi},$$

and evidently $\|u_0\|_{L^2(\mathbb{R})} = \sqrt{2}$. Applying the second conclusion of Lemma F.6, we thus get

$$\left\| u_{\text{rough}} - \frac{2}{\sqrt{\pi}}\cos(\tfrac{\pi}{2}(\,\cdot\,))\mathbb{1}_{[-1,1]}\right\|_{L^2(\mathbb{R})} \le \frac{\sqrt{\pi}}{3^{T_{\text{rough}}-1}} \tag{F.43}$$

as long as $T_{\text{rough}} \ge 2$.

Rough alignment stage. To establish progress by the iteration (F.41), we combine a standard optimization analysis under a lower bound on the magnitude of the gradient with Lemma F.9, which gives a lower bound on the 'nominal' value of the gradient, and a basic perturbation analysis that uses the control we have established in the previous step between u_{rough} and its nominal value.

First, by the fact that $T_{\text{rough}} \ge 2$ and $\sigma \le \frac{1}{100}$, we can apply Lemma F.10. We perform a landscape analysis of the loss $\mathcal{L}^\sigma(\,\cdot\,, u_{\text{rough}})$, where we relate it to properties of the 'nominal loss' $\mathcal{L}^\sigma(\,\cdot\,, \bar{u}_{\text{rough}})$, to guarantee progress of the gradient iteration (F.41). The initialization $\nu_0 \sim \text{Unif}([0, 2\pi])$, and following the proof of Lemma F.9, we see that the objective $\nu \mapsto \mathcal{L}^\sigma(\nu, u)$ (for any σ and any u) is $\pi/2$-periodic and has reflection symmetry about $\nu_\natural$ on the interval $[\nu_\natural - \pi/4, \nu_\natural + \pi/4]$. This implies that the landscape (and hence the behavior of the gradient descent iterates) is determined for all ν by its behavior on the domain $[\nu_\natural - \pi/4, \nu_\natural + \pi/4]$, and we can therefore assume that $\nu_0 \in [\nu_\natural - \pi/4, \nu_\natural + \pi/4]$; it then follows by the uniform initialization that with probability at least $(\pi/7)/(\pi/4) = 4/7$, we have

$$|\nu_0 - \nu_\natural| \le \pi/7.$$

This means we can invoke the lower bound in Lemma F.9 to obtain that

$$\begin{aligned}\text{sign}(\nu_0 - \nu_\natural)\cdot\nabla_\nu\mathcal{L}^\sigma(\nu_0, \bar{u}_{\text{rough}}) &\gtrsim \sin(|\nu_0 - \nu_\natural|)\\ &\gtrsim |\nu_0 - \nu_\natural|,\end{aligned}$$

where the last inequality uses that $\sin x \ge (2/\pi)x$ when $0 \le x \le \pi/2$. Meanwhile, by the first estimate of the second assertion in Lemma F.10 and (F.43), this implies

$$\text{sign}(\nu_0 - \nu_\natural)\cdot\nabla_\nu\mathcal{L}^\sigma(\nu_0, u_{\text{rough}}) \gtrsim |\nu_0 - \nu_\natural| - \frac{3^{-T_{\text{rough}}}}{\sigma^2}. \tag{F.44}$$

Next, using the upper bound in Lemma F.9, we have that for any ν,

$$|\nabla_\nu\mathcal{L}^\sigma(\nu, \bar{u}_{\text{rough}})| \lesssim |\nu - \nu_\natural|.$$

Combining this with the first estimate of the second assertion in Lemma F.10 and (F.43), as above, we obtain

$$|\nabla_\nu\mathcal{L}^\sigma(\nu, u_{\text{rough}})| \lesssim |\nu - \nu_\natural| + \frac{3^{-T_{\text{rough}}}}{\sigma^2}.$$

In particular, choosing $\beta \le c$ for an absolute constant $c \le 1$ and $T_{\text{rough}} \gtrsim -\log(\sigma^2\varepsilon)$ for any $\varepsilon > 0$, we have

$$\beta|\nabla_\nu\mathcal{L}^\sigma(\nu, u_{\text{rough}})| \le |\nu - \nu_\natural| + \varepsilon. \tag{F.45}$$

Now, for $\gamma > 0$, define the domains

$$S_\gamma = \{\nu \in \mathbb{R} \mid |\nu - \nu_\natural| \ge \gamma\}.$$

Fix $0 < \varepsilon < \pi/14$. We are going to argue that after T_ν iterations of (F.41), the last iterate $\hat{\nu}$ satisfies $\hat{\nu} \in S^{\mathsf{c}}_{3\varepsilon}$. We start by proving two invariants of the sequence of iterates (F.41). First, note that

$$\begin{aligned} |\nu_{k+1} - \nu_\natural| &= |\nu_k - \beta \nabla_\nu \mathcal{L}^\sigma(\nu_k, u_{\text{rough}}) - \nu_\natural| \\ &\leq 2|\nu_k - \nu_\natural| + \varepsilon, \end{aligned} \tag{F.46}$$

by the triangle inequality and (F.45). Next, suppose that for some k, we have $\nu_k \in S_\varepsilon \cap S^{\mathsf{c}}_{\pi/7}$. By (F.44), if $T_{\text{rough}} \gtrsim -\log(\sigma^2 \varepsilon)$, we have from (F.44) (via Lemma F.9)

$$\operatorname{sign}(\nu_k - \nu_\natural) \cdot \nabla_\nu \mathcal{L}^\sigma(\nu_k, u_{\text{rough}}) \gtrsim |\nu_k - \nu_\natural|. \tag{F.47}$$

Suppose first that $\nu_k - \nu_\natural \geq 0$. Then (F.47) becomes

$$\nabla_\nu \mathcal{L}^\sigma(\nu_k, u_{\text{rough}}) \gtrsim \nu_k - \nu_\natural.$$

In particular, the gradient at ν_k is nonnegative. This implies

$$\begin{aligned} \nu_{k+1} - \nu_\natural &= \nu_k - \beta \nabla_\nu \mathcal{L}^\sigma(\nu_k, u_{\text{rough}}) - \nu_\natural \\ &\leq (1 - c_0 \beta)(\nu_k - \nu_\natural) \\ &< \nu_k - \nu_\natural, \end{aligned}$$

since $\beta \leq 1$, where $c_0 > 0$ is an absolute constant that we may assume is no larger than $\frac{1}{2}$. We also have, by (F.45), that

$$\nu_{k+1} - \nu_\natural = \nu_k - \beta \nabla_\nu \mathcal{L}^\sigma(\nu_k, u_{\text{rough}}) - \nu_\natural \geq -\varepsilon.$$

But $\nu_k \in S_\varepsilon$, so $\varepsilon \leq |\nu_k - \nu_\natural|$. We conclude

$$|\nu_{k+1} - \nu_\natural| \leq \max\{\varepsilon, (1 - c_0\beta)|\nu_k - \nu_\natural|\} \leq |\nu_k - \nu_\natural|, \tag{F.48}$$

and a completely analogous argument implies the same conclusion in the case where $\nu_k - \nu_\natural \leq 0$. As a consequence, suppose now that for some k, we have $\nu_k \in S^{\mathsf{c}}_{3\varepsilon}$. If in fact $\nu_k \in S^{\mathsf{c}}_{\varepsilon}$, we know immediately from (F.46) that $\nu_{k+1} \in S^{\mathsf{c}}_{3\varepsilon}$. On the other hand, if instead $\nu_k \in S_\varepsilon \cap S^{\mathsf{c}}_{3\varepsilon}$, we have immediately from (F.48) that $\nu_{k+1} \in S^{\mathsf{c}}_{3\varepsilon}$. We conclude the full non-escape invariant:

$$\nu_k \in S^{\mathsf{c}}_{3\varepsilon} \implies \nu_{k+1} \in S^{\mathsf{c}}_{3\varepsilon}. \tag{F.49}$$

We can now give an inductive argument to obtain the desired convergence, namely that

$$\hat{\nu} \in S^{\mathsf{c}}_{3\varepsilon}.$$

First, if $\nu_0 \in S^{\mathsf{c}}_{3\varepsilon}$, we are done immediately, by (F.49). If not, then by the preceding parameter choices and assumption on the initialization we have $\nu_0 \in S_{3\varepsilon} \cap S^{\mathsf{c}}_{\pi/7}$ and therefore $\nu_0 \in S_\varepsilon \cap S^{\mathsf{c}}_{\pi/7}$, so that by (F.48), we obtain

$$\nu_1 \in S^{\mathsf{c}}_{\max\{\varepsilon, (1-c_0\beta)|\nu_0 - \nu_\natural|\}}.$$

At this point, notice that by assumption $(1 - c_0\beta)|\nu_0 - \nu_\natural| \geq \frac{3}{2}\varepsilon$, so in fact

$$\nu_1 \in S^{\mathsf{c}}_{(1-c_0\beta)|\nu_0 - \nu_\natural|}.$$

Proceeding inductively in this way, it follows that at iteration $k \in \mathbb{N}$ we either have $\nu_k \in S^{\mathsf{c}}_{3\varepsilon}$ or that $\nu_{k-1} \in S_{3\varepsilon} \cap S^{\mathsf{c}}_{\pi/7}$ and

$$\nu_k \in S^{\mathsf{c}}_{(1-c_0\beta)|\nu_{k-1} - \nu_\natural|}.$$

Unraveling this recurrence gives

$$\nu_k \in S^{\mathsf{c}}_{(1-c_0\beta)^k|\nu_0 - \nu_\natural|}.$$

Thus, as soon as

$$T_\nu \gtrsim \frac{\log(3\varepsilon)}{\log(1 - c_0\beta)},$$

we have $\hat{\nu} \in S^{\mathsf{c}}_{3\varepsilon}$, i.e. that

$$|\hat{\nu} - \nu_\natural| \leq 3\varepsilon. \tag{F.50}$$

Refined factorization stage. In this stage, we run power method to refine the factorization u_{rough}, using the fact that $\hat{\nu} \approx \nu_\natural$ to argue that the relevant operator for the refinement power method (F.42) (c.f. (F.36) and (F.37)), namely

$$\hat{X} = \frac{1}{2}\left(X_\natural \circ \boldsymbol{\tau}_{\hat{\nu}-\nu_\natural} + X_\natural \circ \boldsymbol{\tau}_{-(\hat{\nu}-\nu_\natural)}\right),$$

is sufficiently close to $X_\natural$ that we can guarantee the progress of the power method (F.42) by applying spectral properties of $X_\natural = u_\natural u_\natural^*$ in Lemma F.6. To do this, we need to prove that the spectrum of $\hat{X}$ has a gap. First, we note that $\hat{X}$ is Hilbert-Schmidt, by the triangle inequality and the fact that it is a Fredholm operator (together with the fact that $f \mapsto f \circ \boldsymbol{\tau}_\nu$ is a unitary transformation of $L^2(\mathbb{R}^2)$):

$$\begin{aligned}\|\hat{X}\|_{\text{HS}} &\leq \frac{1}{2}\left(\left\|X_\natural \circ \boldsymbol{\tau}_{\hat{\nu}-\nu_\natural}\right\|_{\text{HS}} + \|X_\natural \circ \boldsymbol{\tau}_{-(\hat{\nu}-\nu_\natural)}\|_{\text{HS}}\right) \\ &= \|X_\natural\|_{\text{HS}}. \end{aligned} \tag{F.51}$$

This means that $\hat{X}$ is a compact operator; we have therefore $\hat{X} = \sum_{i=1}^{\infty} \lambda_i v_i v_i^*$ for eigenvalues $(\lambda_i)_{i\in\mathbb{N}} \subset \mathbb{R}$ and an orthonormal basis of eigenfunctions $(v_i)_{i\in\mathbb{N}}$ (c.f. [12, §B]). Without loss of generality, we assume that the sequence is ordered such that $|\lambda_1| = \|\hat{X}\|$. To show that the spectrum has a gap in the way that is needed to apply Lemma F.6, we need to show that $\lambda_1 > 0$ and that the rest of the spectrum is bounded in magnitude away from λ_1. We will establish the latter first. Note that

$$\begin{aligned}\sup_{i\geq 2}|\lambda_i|^2 \leq \sum_{i=2}^{\infty}|\lambda_i|^2 &= -|\lambda_1|^2 + \sum_{i=1}^{\infty}|\lambda_i|^2 \\ &= \|\hat{X}\|_{\text{HS}}^2 - \|\hat{X}\|^2.\end{aligned}$$

By the triangle inequality and (as above) the fact that $\|\cdot\| \leq \|\cdot\|_{\text{HS}}$, we have

$$\begin{aligned}\|\hat{X}\| &\geq \|X_\natural\| - \|X_\natural - \hat{X}\| \\ &\geq \|X_\natural\| - \|X_\natural - \hat{X}\|_{\text{HS}} \\ &= \|X_\natural\|_{\text{HS}} - \|X_\natural - \hat{X}\|_{\text{HS}}\end{aligned}$$

where we used the fact that $X_\natural = u_\natural u_\natural^*$, so that $\|X_\natural\|_{\text{HS}} = u_\natural^* u_\natural$ coincides with $\|X_\natural\| = \sup_{\|f\|_{L^2}\leq 1}|\langle u_\natural, f\rangle|\|u_\natural\|_{L^2} = \|u_\natural\|_{L^2}^2$ (apply the Schwarz inequality). Thus, if $\|X_\natural - \hat{X}\|_{\text{HS}} \leq \|X_\natural\|_{\text{HS}}$, we have

$$\begin{aligned}\sup_{i\geq 2}|\lambda_i|^2 &\leq \|\hat{X}\|_{\text{HS}}^2 - \left(\|X_\natural\|_{\text{HS}} - \|X_\natural - \hat{X}\|_{\text{HS}}\right)^2 \\ &\leq 2\|X_\natural\|_{\text{HS}}\|X_\natural - \hat{X}\|_{\text{HS}},\end{aligned}$$

where the second inequality applies (F.51) and discards the (negative) second-order term. Since $u_\natural = \mathbb{1}_{[-1/\sqrt{2},1/\sqrt{2}]}$, we have $\|X_\natural\|_{\text{HS}} = \sqrt{2}$, and to proceed using the above it suffices to control $\|X_\natural - \hat{X}\|_{\text{HS}}$ and show that it is no larger than $\sqrt{2}$. To this end, we have

$$\begin{aligned}\left\|\hat{X} - X_\natural\right\| &\leq \left\|\hat{X} - X_\natural\right\|_{\text{HS}} \\ &\leq \frac{1}{2}\left\|X_\natural \circ \boldsymbol{\tau}_{\hat{\nu}-\nu_\natural} - X_\natural\right\|_{\text{HS}} + \frac{1}{2}\left\|X_\natural \circ \boldsymbol{\tau}_{-(\hat{\nu}-\nu_\natural)} - X_\natural\right\|_{\text{HS}},\end{aligned}$$

where the second line uses the triangle inequality. Below, write $\alpha = 1/\sqrt{2}$. One has for any $\varepsilon > 0$

$$\begin{aligned}\|X_\natural \circ \boldsymbol{\tau}_\varepsilon - X_\natural\|_{\text{HS}}^2 &= 2\|X_\natural\|_{\text{HS}}^2 - 2\langle X_\natural \circ \boldsymbol{\tau}_\varepsilon, X_\natural\rangle_{L^2(\mathbb{R}^2)} \\ &= 8\alpha^2 - 2\langle X_\natural \circ \boldsymbol{\tau}_\varepsilon, X_\natural\rangle_{L^2(\mathbb{R}^2)},\end{aligned}$$

because the operators are Fredholm operators and $\boldsymbol{\tau}_\varepsilon$ is a unitary transformation. To estimate the cross term, we argue geometrically. For $|\varepsilon| \leq 1$, we have the estimate $|\cos\varepsilon| + |\sin\varepsilon| \leq 1 + |\varepsilon|$.This implies that

$$\begin{aligned}\|\boldsymbol{R}_\varepsilon \boldsymbol{x}\|_\infty &\leq \|\boldsymbol{R}_\varepsilon\|_{\infty\to\infty}\|\boldsymbol{x}\|_\infty \\ &\leq \|\boldsymbol{x}\|_\infty\left(|\cos\varepsilon| + |\sin\varepsilon|\right) \\ &\leq \|\boldsymbol{x}\|_\infty(1 + |\varepsilon|).\end{aligned}$$

Points $\boldsymbol{x} \in \mathbb{R}^2$ where the inner product $\langle X_\natural \circ \boldsymbol{\tau}_\varepsilon, X_\natural \rangle_{L^2(\mathbb{R}^2)}$ is positive are those where $\|\boldsymbol{x}\|_\infty \leq \alpha$ and $\|\boldsymbol{R}_\varepsilon \boldsymbol{x}\|_\infty \leq \alpha$. By the previous estimate, both conditions occur when $\|\boldsymbol{x}\|_\infty \leq \alpha/(1+|\varepsilon|)$. Since $1/(1+|\varepsilon|)^2 \geq (1-|\varepsilon|)^2$ if $|\varepsilon| \leq 1$, this implies

$$\begin{aligned}\langle X_\natural \circ \boldsymbol{\tau}_\varepsilon, X_\natural \rangle_{L^2(\mathbb{R}^2)} &\geq \int_{\|\boldsymbol{x}\|_\infty \leq \alpha/(1+|\varepsilon|)} \mathrm{d}\boldsymbol{x} \\ &= \frac{4\alpha^2}{(1+|\varepsilon|)^2} \\ &\geq 4\alpha^2 - 8\alpha^2|\varepsilon|,\end{aligned}$$

so

$$\|X_\natural \circ \boldsymbol{\tau}_\varepsilon - X_\natural\|_{\mathrm{HS}}^2 \leq 16\alpha^2|\varepsilon|,$$

and thus, since $|\nu_\natural - \hat{\nu}| \leq 1$ by the reduction-by-symmetry to the domain $[\nu_\natural - \pi/4, \nu_\natural + \pi/4]$ given in the previous phase of the argument,

$$\|\hat{X} - X_\natural\| \leq \|\hat{X} - X_\natural\|_{\mathrm{HS}} \leq 2\sqrt{2}\sqrt{|\hat{\nu} - \nu_\natural|}. \tag{F.52}$$

In particular, if $|\hat{\nu} - \nu_\natural| \leq \frac{1}{4}$, we have $\|\hat{X} - X_\natural\|_{\mathrm{HS}} \leq \|X_\natural\|_{\mathrm{HS}}$, and by the above

$$\sup_{i \geq 2} |\lambda_i| \leq 2\sqrt{2}|\nu_\natural - \hat{\nu}|^{1/4}.$$

Meanwhile, under this condition the above estimates yield

$$\begin{aligned}|\lambda_1| = \|\hat{X}\| &\leq \|X_\natural\| + \|X_\natural - \hat{X}\| \\ &\leq \sqrt{2} + 2\sqrt{2}|\nu_\natural - \hat{\nu}|^{1/2},\end{aligned} \tag{F.53}$$

and moreover, by the Schwarz inequality,

$$\begin{aligned}\langle u_\natural, \hat{X}[u_\natural] \rangle_{L^2(\mathbb{R})} &\geq \langle u_\natural, X_\natural[u_\natural] \rangle - \|u_\natural\|_{L^2(\mathbb{R})}^2 \|\hat{X} - X_\natural\| \\ &\geq \|u_\natural\|_{L^2(\mathbb{R})}^4 - 2\sqrt{2}\|u_\natural\|_{L^2(\mathbb{R})}^2 |\nu_\natural - \hat{\nu}|^{1/2}\end{aligned}$$

Since $\|u_\natural\|_{L^2(\mathbb{R})} = 2^{1/4}$, this means that if $|\hat{\nu} - \nu_\natural| \leq \frac{1}{64}$, we have

$$\left\langle \frac{u_\natural}{\|u_\natural\|_{L^2}}, \hat{X}\left[\frac{u_\natural}{\|u_\natural\|_{L^2}}\right] \right\rangle_{L^2(\mathbb{R})} \geq \frac{3\sqrt{2}}{4},$$

and since, under this condition, we have

$$\sup_{i \geq 2} |\lambda_i| \leq 1,$$

we conclude that

$$\max_{i \in \mathbb{N}} \lambda_i = \max_{\|u\|_{L^2(\mathbb{R})} \leq 1} \langle u, \hat{X}[u] \rangle_{L^2(\mathbb{R})} \geq \left\langle \frac{u_\natural}{\|u_\natural\|_{L^2}}, \hat{X}\left[\frac{u_\natural}{\|u_\natural\|_{L^2}}\right] \right\rangle_{L^2(\mathbb{R})} > \sup_{i \geq 2} |\lambda_i|.$$

(c.f. the proof of Lemma F.19). In particular, we have $\lambda_1 > 0$. Following as well the above arguments, we have

$$|\lambda_1| \geq \sqrt{2} - 2\sqrt{2}|\nu_\natural - \hat{\nu}|^{1/2} \geq \frac{3\sqrt{2}}{4}, \tag{F.54}$$

which implies the gap condition

$$\frac{\sup_{i \geq 2} |\lambda_i|}{\lambda_1} \leq \frac{2|\nu_\natural - \hat{\nu}|^{1/4}}{1 - 2|\nu_\natural - \hat{\nu}|^{1/2}} \leq \frac{8}{3}|\nu_\natural - \hat{\nu}|^{1/4} < 1 \tag{F.55}$$

under the preceding assumptions. Using these characterizations of the spectral gap, we can conveniently also apply [3, Proposition 6.1] to obtain

$$\left\| \frac{u_\natural}{\|u_\natural\|_{L^2}} \frac{u_\natural^*}{\|u_\natural\|_{L^2}} - v_1 v_1^* \right\|_{\mathrm{HS}} \le \frac{2}{\sqrt{2}-1} \|X_\natural - \hat{X}\|_{\mathrm{HS}} \le 14\sqrt{|\nu_\natural - \hat{\nu}|}.$$

Meanwhile, we have

$$\left\| \frac{u_\natural}{\|u_\natural\|_{L^2}} \frac{u_\natural^*}{\|u_\natural\|_{L^2}} - v_1 v_1^* \right\|_{\mathrm{HS}}^2 = 2\left(1 - \left| \left\langle \frac{u_\natural}{\|u_\natural\|_{L^2}}, v_1 \right\rangle \right|^2\right).$$

Since v_1 is only defined up to sign, let us suppose without loss of generality that v_1 is such that $\langle v_1, u_\natural \rangle_{L^2} \ge 0$. Proceeding, we then obtain

$$\begin{aligned}
\left\| \frac{u_\natural}{\|u_\natural\|_{L^2}} \frac{u_\natural^*}{\|u_\natural\|_{L^2}} - v_1 v_1^* \right\|_{\mathrm{HS}}^2 &= 2\left(1 - \left| \left\langle \frac{u_\natural}{\|u_\natural\|_{L^2}}, v_1 \right\rangle \right|\right)\left(1 + \left| \left\langle \frac{u_\natural}{\|u_\natural\|_{L^2}}, v_1 \right\rangle \right|\right) \\
&\ge 2\left(1 - \left| \left\langle \frac{u_\natural}{\|u_\natural\|_{L^2}}, v_1 \right\rangle \right|\right) \\
&= \left\| \frac{u_\natural}{\|u_\natural\|_{L^2}} - v_1 \right\|_{L^2(\mathbb{R})}^2.
\end{aligned}$$

Combining, this gives

$$\left\| 2^{-1/4} u_\natural - v_1 \right\|_{L^2(\mathbb{R})} \le 14\sqrt{|\nu_\natural - \hat{\nu}|}. \tag{F.56}$$

This allows us to lower bound the correlation between the power method initialization u_{rough} and the target eigenvector: by the triangle inequality and the Schwarz inequality,

$$\begin{aligned}
\langle v_1, u_{\mathrm{rough}} \rangle &\ge \langle v_1, \bar{u}_{\mathrm{rough}} \rangle - \|u_{\mathrm{rough}} - \bar{u}_{\mathrm{rough}}\|_{L^2(\mathbb{R})} \\
&\ge 2^{-1/4} \langle u_\natural, \bar{u}_{\mathrm{rough}} \rangle - \|v_1 - 2^{-1/4} u_\natural\|_{L^2(\mathbb{R})} - \|u_{\mathrm{rough}} - \bar{u}_{\mathrm{rough}}\|_{L^2(\mathbb{R})}.
\end{aligned}$$

We calculate

$$2^{-1/4} \langle u_\natural, \bar{u}_{\mathrm{rough}} \rangle = \frac{2}{2^{1/4}\sqrt{\pi}} \int_{-1}^{1} \cos(\pi x/2)\, \mathrm{d}x = \frac{8}{2^{1/4}\pi^{3/2}},$$

so that, by (F.43) and (F.56), we have for $T_{\mathrm{rough}} \ge 4$ and $|\nu_\natural - \hat{\nu}| \le \frac{1}{256}$ that

$$\langle v_1, u_{\mathrm{rough}} \rangle \ge \frac{1}{4}. \tag{F.57}$$

Moreover, it allows us to control the unnormalized distance between the power method output and the target in terms of the normalized distance: the triangle inequality gives

$$\begin{aligned}
\left\| u_\natural - \sqrt{\lambda_1} v_1 \right\|_{L^2(\mathbb{R})} &\le \max\left\{ \|u_\natural\|_{L^2(\mathbb{R})}, \sqrt{\lambda_1} \right\} \left\| \frac{u_\natural}{\|u_\natural\|_{L^2}} - v_1 \right\|_{L^2(\mathbb{R})} + \left| \|u_\natural\|_{L^2(\mathbb{R})} - \sqrt{\lambda_1} \right| \\
&\le 14\left(2^{1/4} + 2^{3/8}|\nu_\natural - \hat{\nu}|^{1/4}\right)\sqrt{|\nu_\natural - \hat{\nu}|} + \left| \|u_\natural\|_{L^2(\mathbb{R})} - \sqrt{\lambda_1} \right|,
\end{aligned}$$

where the second line applies (F.53). Meanwhile, (F.54) and (F.53) imply

$$2^{1/4}\left(\sqrt{1 - 2|\nu_\natural - \hat{\nu}|^{1/2}} - 1\right) \le \sqrt{\lambda_1} - 2^{1/4} \le 2^{1/4}\left(\sqrt{1 + 2|\nu_\natural - \hat{\nu}|^{1/2}} - 1\right),$$

and given that $2|\nu_\natural - \hat{\nu}|^{1/2} \le \frac{1}{8}$ by our assumptions, the inequalities $\sqrt{1+x} \le 1 + x/2$ and $\sqrt{1-x} \ge 1 - x$ (the latter valid if $0 \le x \le 1$) lead to the bounds

$$-2^{5/4}|\nu_\natural - \hat{\nu}|^{1/2} \le \sqrt{\lambda_1} - 2^{1/4} \le 2^{5/4}|\nu_\natural - \hat{\nu}|^{1/2},$$

so, plugging into the previous estimate, we obtain

$$\left\| u_\natural - \sqrt{\lambda_1} v_1 \right\|_{L^2(\mathbb{R})} \leq 30 \sqrt{|\nu_\natural - \hat{\nu}|} \tag{F.58}$$

after worst-casing constants. We can finally apply Lemma F.6 with the properties (F.55) and (F.57) together with the triangle inequality and (F.58) to obtain that the iteration (F.42) satisfies

$$\|\hat{u} - u_\natural\|_{L^2(\mathbb{R})} \leq 30 \sqrt{|\nu_\natural - \hat{\nu}|} + 28 \left(\frac{8}{3} |\nu_\natural - \hat{\nu}|^{1/4} \right)^{T_u}$$

after worst-casing constants slightly.

Concluding the result. Finally, we instantiate our results above with appropriate parameter choices to obtain the desired conclusion. We have shown the following: there are absolute constants c_1, C_1, $C_2 > 0$ such that for any parameters σ, β satisfying

$$\begin{aligned} \sigma^2 &\leq \tfrac{1}{10^4}, \\ \beta &\leq c_1, \end{aligned}$$

for any $0 < \varepsilon \leq \frac{1}{768}$, if the iteration counts satisfy

$$\begin{aligned} T_{\text{rough}} &\geq -C_1 \log(\sigma^2 \varepsilon), \\ T_\nu &\geq -\frac{C_2 \log(3\varepsilon)}{\beta}, \\ T_u &\geq 16, \end{aligned}$$

then with probability over the random initialization of ν_0 at least $4/7$, one has

$$\begin{aligned} |\hat{\nu} - \nu_\natural| &\leq 3\varepsilon, \\ \|\hat{u} - u_\natural\|_{L^2(\mathbb{R})} &\leq 31\sqrt{\varepsilon}. \end{aligned}$$

The condition on T_u is quite mild because the rate of convergence improves with the quality of the output of the alignment stage of the algorithm, since the target $u_\natural u_\natural^*$ is rank one. The condition on T_ν above takes advantage of the fact that when $0 \leq x \leq \frac{1}{2}$, we have by concavity $\log(1-x) \geq (-2\log 2)x$ to simplify the stated bound in our previous work.

□

Remark F.2. Theorem F.2 establishes a linear rate of convergence of both the alignment iterates ν_k and the representation iterations that generate u_{rough} and $\hat{u}$ to the true parameters of the template, up to symmetry. The dependence on the other problem parameters in these rates, namely the smoothing σ^2 and the step size β, is about as mild as one would hope for: the smoothing level σ^2 only enters the rates logarithmically, and the step size is only required to be smaller than an absolute constant, which is reflected as a linear dependence in the rate of convergence of the alignment step of the algorithm. The issue of smoothing represents an interesting conceptual takeaway from our analysis, with regards to modern 3D representation approaches like TILTED which do not explicitly incorporate a classical coarse-to-fine smoothing schedule, as in, for example, image registration [9]. Our proofs demonstrate that the reason smoothing is not necessary for precise local convergence is that *computational constraints on the representation capacity of the method (here, a rank-one matrix) and the L^2 loss create texture gradients (i.e., blurry images) when optimizing the representation*, and these texture gradients cause the subsequent alignment landscape to be smoother than one might otherwise expect. For example, when σ^2 is small, the smoothed template $\varphi_{\sigma^2} * X_\natural$ has a Lipschitz constant on the order of $1/\sigma$, due to sharp edges in $X_\natural$; nevertheless, this sharpness does not reflect in our rates as a consequence of the blessings of capacity-constrained inexact representation.

Theorem F.2 contains a hypothesis on the probability of success, which is asserted to be at least $\frac{4}{7}$; since this value is larger than $\frac{1}{2}$, it is theoretically possible to boost the success probability to an arbitrarily-high level by running multiple independent trials of the algorithm and aggregating the outputs appropriately. In practice, of course, the algorithm succeeds with probability one on the template $X_\natural$: the discrepancy is due to the technical need to prove that the nonconvex alignment

landscape $\nu \mapsto \mathcal{L}^\sigma(\nu, u_{\text{rough}})$ has suitable negative curvature in neighborhoods of the maximizers $(\nu_\natural + \frac{\pi}{4}) + \frac{\pi}{2}\mathbb{Z}$. This "benign global geometry", in the sense of Zhang *et al.* [22], has been studied in other contexts [18, 20, 29], although the mixture of discrete and continuous symmetries in the (ν, u) landscape of (F.34) is somewhat distinguished. In general, we have endeavored to keep the optimization analysis in Theorem F.2 as elementary as possible, and we have not made attempts to optimize absolute constants. Simple and standard modifications to the proofs can be made to yield slightly better constants and rates, at the cost of additional technicality [10, 18, 22].

Remark F.3. We discuss three directions of extension for Theorem F.2 below.

1. **Joint factorization and alignment.** The algorithm we study, in its use of the power method as a standin for matrix factorization as in (F.34) as well as its use of "block" alternating minimization iterations rather than alternating gradient steps on ν and u, differs from our implementation of TILTED in ways that present important directions for further technical improvement. It seems to us that extending our analysis to the case of alternating gradient steps would require some additional conceptual insight (e.g., the identification of a conserved quantity): various technical components of the current alignment argument are delicate and break when the factorization target is not constant. A similar issue is associated with the extension of the result to $\nu_\natural \neq \pi/4$, although this case seems "easier"; e.g., perturbative analogues of Lemma F.1 for other values of $\nu_\natural$ would suffice here. In the alternating (ν, u) setting, there is also a challenge associated with the fact that texture gradients induced by inexact factorization become smaller as the alignment becomes more accurate, making the landscape nonsmooth. Rather than introducing auxiliary smoothing in this setting, it may be most relevant to practice to study the nonsmooth landscape directly, à la [19]; the analysis can be less technical in our setting since there is no statistical component to the problem. Extending the loss (F.34) to the setting of overparameterized matrix factorization is also interesting; developing this extension in the context of observations $X_\natural$ with background clutter, as discussed below, may be the most natural setting.

2. **Extensions to multi-object scenes and three dimensions.** Establishing a direct 3D analogue of Theorem F.2 seems to be mostly technical. Extending the result to apply to scenes $X_\natural$ with other objects and background clutter present seems more challenging: the proof of Theorem F.2 presents a perturbative framework for analyzing TILTED that should not be hard to extend to 'perturbed' observations $X_\natural$ when the magnitude of the perturbation is small, but getting insights into how the algorithm can be changed to cope with the kinds of structured perturbations that arise in real-world scenes (e.g., a scene with objects with shape content that cannot be axis-aligned as well as the square, like people, but that nonetheless contains enough 'prominent' axis-alignable components, such as buildings and roads in a built environment, for success to be possible) seems to require novel ideas. One path forward here could be to introduce additional appearance components to the square $X_\natural$, such as a texture, and study conditions under which these are sufficiently decorrelated with the shape components of the template for success to remain possible. Another possibility is to study the overparameterized case and separate distinct factorization components into "groups", as in our implementation of TILTED, which have their own transformations and can thus represent distinct parts of the scene; the necessary symmetry-breaking aspects of such a result feel reminiscent of analyses of dictionary learning [22], but feel significantly more challenging due to the need to localize different objects via factorization in the setting of TILTED.

3. **Computing with a MLP.** An extremely important avenue for extension of Theorem F.2 is to go beyond the linear representation studied there and introduce a neural network for representing the scene. This presents an additional challenge with respect to disentangling appearance and shape versus our current analysis: there, the rank-one capacity constraint on the factorization leads to inexact intermediate factorizations that create texture gradients to help alignment, and alignment improvements help to further improve the representation. With a MLP, there seems to be some natural capacity to represent coordinate rotations (c.f. Section F.4)—understanding how initialization and implicit bias of gradient descent training preserves the disentangled learning of appearance and alignment that we prove occurs in the linear model is a fascinating direction for future work.

F.2.1 Supporting Results

Lemma F.6. *Let $\mathcal{T} : L^2(\mathbb{R}) \to L^2(\mathbb{R})$ be a nonzero self-adjoint Hilbert-Schmidt operator with corresponding eigenvalues $(\lambda_k)_{k\in\mathbb{N}}$ and orthonormal basis $(v_k)_{k\in\mathbb{N}} \subset L^2(\mathbb{R})$. Without loss of generality, suppose that $\|\mathcal{T}\| = |\lambda_1| > 0$. Suppose moreover that $\lambda_1 > 0$,[4] and that the spectrum has a gap, i.e., that for some $0 < \gamma \leq 1$ we have $\lambda_1 - |\lambda_k| \geq \gamma\lambda_1$ for all $k > 1$.*

[4] If $\lambda_1 < 0$, it is necessary to take absolute values in order to obtain the result asserted here: consider for example the case where $\mathcal{T} = -\text{Id}$, so that $u_k = (-1)^k u_0$ if $\|u_0\|_{L^2(\mathbb{R})} = 1$.

Consider power method on $\mathcal{T}$, starting from initialization $u_0 \in L^2(\mathbb{R})$:

$$u_{k+1} = \frac{\mathcal{T}u_k}{\|\mathcal{T}u_k\|_{L^2(\mathbb{R})}}.$$

Then if $|\langle v_1, u_0\rangle_{L^2(\mathbb{R})}| \geq \eta > 0$, it holds

$$\left\|u_k - \operatorname{sign}(\langle v_1, u_0\rangle_{L^2(\mathbb{R})})v_1\right\|^2_{L^2(\mathbb{R})} \leq 2(1-\gamma)^{2k}\frac{\|u_0\|^2_{L^2(\mathbb{R})}}{\eta^2}.$$

In addition, if the iteration count satisfies

$$k \geq \frac{\log(\eta/4\|u_0\|_{L^2(\mathbb{R})})}{\log(1-\gamma)},$$

then the 'rank-one approximating factor' error satisfies

$$\left\|\sqrt{u_k^*\mathcal{T}u_k}u_k - \sqrt{\lambda_1}\operatorname{sign}(\langle v_1, u_0\rangle_{L^2(\mathbb{R})})v_1\right\|^2_{L^2(\mathbb{R})} \leq 18\lambda_1(1-\gamma)^{2k}\frac{\|u_0\|^2_{L^2(\mathbb{R})}}{\eta^2}.$$

Proof. We apply the standard argument; under the assumption of a Hilbert-Schmidt operator with a gapped spectrum, as we have made here, the standard argument's convergence is actually dimension-free, in contrast to the general case (c.f. [6]). We use basic notions from the analysis of self-adjoint Hilbert-Schmidt operators in the proof (see [11, §B]).

The main observation to make is that the update equation for u_{k+1} in the definition of the power method is a 0-absolutely-homogeneous function of u_k. This implies

$$u_k = \frac{\mathcal{T}^k u_0}{\|\mathcal{T}^k u_0\|_{L^2(\mathbb{R})}}.$$

It is clear that this iteration is well-defined, i.e., that for every k one has $\mathcal{T}^k u_0 \neq 0$, by the assumption that $|\langle v_1, u_0\rangle| > 0$ (and the fact that $|\lambda_1| > 0$), since, as we will use below,

$$\mathcal{T}^k u_0 = \sum_{l=1}^{\infty}\lambda_l^k\langle v_l, u_0\rangle_{L^2(\mathbb{R})}v_l,$$
$$\left\|\mathcal{T}^k u_0\right\|^2_{L^2(\mathbb{R})} = \sum_{l=1}^{\infty}\lambda_l^{2k}|\langle v_l, u_0\rangle_{L^2(\mathbb{R})}|^2,$$

since the sequence (v_l) is an orthonormal basis. Moreover, notice that if we initialize the power method with $-u_0$ instead of u_0, we end up only changing the sign of the output u_k; hence we can assume below without loss of generality that $\langle v_1, u_0\rangle > 0$. Now, expanding the square shows that

$$\begin{aligned}\|u_k - v_1\|^2_{L^2} &= \left\|\frac{\mathcal{T}^k u_0}{\|\mathcal{T}^k u_0\|_{L^2(\mathbb{R})}} - v_1\right\|^2_{L^2} \\ &= 2\left(1 - \frac{\lambda_1^k\langle v_1, u_0\rangle_{L^2(\mathbb{R})}}{\|\mathcal{T}^k u_0\|_{L^2(\mathbb{R})}}\right).\end{aligned}$$

Since $\lambda_1 > 0$, we have

$$\begin{aligned}\frac{\|\mathcal{T}^k u_0\|_{L^2(\mathbb{R})}}{\lambda_1^k\langle v_1, u_0\rangle_{L^2(\mathbb{R})}} &= \left(\frac{1}{\lambda_1^{2k}\langle v_1, u_0\rangle^2_{L^2(\mathbb{R})}}\sum_{l=1}^{\infty}\lambda_l^{2k}|\langle v_l, u_0\rangle_{L^2(\mathbb{R})}|^2\right)^{1/2} \\ &= \left(1 + \sum_{l=2}^{\infty}\left(\frac{|\lambda_l|}{|\lambda_1|}\right)^{2k}\frac{\langle v_l, u_0\rangle^2_{L^2(\mathbb{R})}}{\langle v_1, u_0\rangle^2_{L^2(\mathbb{R})}}\right)^{1/2}.\end{aligned}$$

The gapped assumption implies that

$$\frac{|\lambda_l|}{|\lambda_1|} \leq 1 - \gamma,$$

and the lower bound $\langle v_1, u_0\rangle^2_{L^2(\mathbb{R})} \geq \eta^2$ then implies

$$\frac{\|\mathcal{T}^k u_0\|_{L^2(\mathbb{R})}}{\lambda_1^k \langle v_1, u_0\rangle_{L^2(\mathbb{R})}} \leq \left(1 + (1-\gamma)^{2k} \frac{\|u_0\|^2_{L^2(\mathbb{R})}}{\eta^2}\right)^{1/2}.$$

For $x \geq 0$, the function $x \mapsto 1 - (1+x)^{-1/2}$ is increasing and concave, and therefore satisfies $1 - (1+x)^{-1/2} \leq \frac{1}{2}x$. As a result, we have

$$\|u_k - v_1\|^2_{L^2} \leq 2(1-\gamma)^{2k} \frac{\|u_0\|^2_{L^2(\mathbb{R})}}{\eta^2},$$

as claimed.

To obtain the claimed estimate for the rank-one approximating factor error, we use the triangle inequality and a bound for the square root when its argument is sufficiently far from 0. Notice that $\sqrt{\lambda_1} = \sqrt{v_1^* \mathcal{T} v_1}$, and by the triangle inequality

$$\begin{aligned} |u_k^* \mathcal{T} u_k - \lambda_1| &\leq |u_k^* \mathcal{T} u_k - u_k^* \mathcal{T} v_1| + |u_k^* \mathcal{T} v_1 - v_1^* \mathcal{T} v_1| \\ &\leq 2\|\mathcal{T}\| \|u_k - v_1\|_{L^2(\mathbb{R})} \\ &= 2\lambda_1 \|u_k - v_1\|_{L^2(\mathbb{R})}. \end{aligned} \tag{F.59}$$

by the Schwarz inequality and the fact that u_k and v_1 are unit norm. This implies

$$u_k^* \mathcal{T} u_k \geq \lambda_1 \left(1 - 2\|u_k - v_1\|_{L^2(\mathbb{R})}\right).$$

As a result, if

$$k \geq \frac{\log(\eta/4\|u_0\|_{L^2(\mathbb{R})})}{\log(1-\gamma)},$$

we have $u_k^* \mathcal{T} u_k > 0$, and square roots can be taken without worry. We have by the triangle inequality

$$\begin{aligned} \left\| \sqrt{u_k^* \mathcal{T} u_k} u_k - \sqrt{\lambda_1} v_1 \right\|_{L^2(\mathbb{R})} &\leq \left\| \sqrt{u_k^* \mathcal{T} u_k} u_k - \sqrt{v_1^* \mathcal{T} v_1} u_k \right\|_{L^2(\mathbb{R})} + \left\| \sqrt{v_1^* \mathcal{T} v_1} u_k - \sqrt{v_1^* \mathcal{T} v_1} v_1 \right\|_{L^2(\mathbb{R})} \\ &\leq \left| \sqrt{u_k^* \mathcal{T} u_k} - \sqrt{v_1^* \mathcal{T} v_1} \right| + \sqrt{\lambda_1} \|u_k - v_1\|_{L^2(\mathbb{R})}, \end{aligned}$$

since u_k and v_1 are unit norm. Now, by the fundamental theorem of calculus, we have for any $x, y \geq 0$

$$\left|\sqrt{x} - \sqrt{y}\right| = \frac{1}{2}\left| \int_y^x z^{-1/2} \, \mathrm{d}z \right| \leq \frac{|x-y|}{2\sqrt{\min\{x,y\}}},$$

and in our setting, we have shown above

$$u_k^* \mathcal{T} u_k \geq \left(1 - \frac{1}{\sqrt{2}}\right) \lambda_1$$

by our choice of k. Since $2(1 - 1/\sqrt{2})^{1/2} \geq 1$, it follows

$$\begin{aligned} \left| \sqrt{u_k^* \mathcal{T} u_k} - \sqrt{v_1^* \mathcal{T} v_1} \right| &\leq \frac{1}{\sqrt{\lambda_1}} |u_k^* \mathcal{T} u_k - v_1^* \mathcal{T} v_1| \\ &\leq 2\sqrt{\lambda_1} \|u_k - v_1\|_{L^2(\mathbb{R})}. \end{aligned}$$

where we used (F.59) in the final line. Consequently, we have shown

$$\left\| \sqrt{u_k^* \mathcal{T} u_k} u_k - \sqrt{\lambda_1} v_1 \right\|_{L^2(\mathbb{R})} \leq 3\sqrt{\lambda_1} \|u_k - v_1\|_{L^2(\mathbb{R})},$$

as claimed.

□

Lemma F.7. *For the objective $\mathcal{L}^\sigma$ defined in (F.32), one has*

$$\begin{aligned}
\nabla_\nu \mathcal{L}^\sigma(\nu, \boldsymbol{U}, \boldsymbol{V}) &= -\left\langle \varphi_{\sigma^2} * \left(\boldsymbol{U}\boldsymbol{V}^* \circ \boldsymbol{\tau}_{\nu_\natural - \nu}\right), \left\langle \nabla_{\boldsymbol{x}}[\varphi_{\sigma^2} * X_\natural], \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} (\,\cdot\,) \right\rangle_{\ell^2} \right\rangle_{L^2(\mathbb{R}^2)}, \\
\nabla_{\boldsymbol{U}} \mathcal{L}^\sigma(\nu, \boldsymbol{U}, \boldsymbol{V}) &= -\left[\varphi_{2\sigma^2} * \left(X_\natural \circ \boldsymbol{\tau}_{\nu - \nu_\natural} - \boldsymbol{U}\boldsymbol{V}^*\right)\right] \boldsymbol{V} \\
\nabla_{\boldsymbol{V}} \mathcal{L}^\sigma(\nu, \boldsymbol{U}, \boldsymbol{V}) &= -\left[\varphi_{2\sigma^2} * \left(X_\natural \circ \boldsymbol{\tau}_{\nu_\natural - \nu} - \boldsymbol{V}\boldsymbol{U}^*\right)\right] \boldsymbol{U}.
\end{aligned}$$

Proof. For the gradients of $\mathcal{L}^\sigma$ with respect to $(\boldsymbol{U}, \boldsymbol{V})$, direct calculation using the chain rule for the Fréchet derivative and the duality of $L^2(\mathbb{R})$ gives

$$\begin{aligned}
\nabla_{\boldsymbol{U}} \mathcal{L}^\sigma(\nu, \boldsymbol{U}, \boldsymbol{V}) &= -\left[\varphi_{2\sigma^2} * \left(X_\natural \circ \boldsymbol{\tau}_{\nu - \nu_\natural} - \boldsymbol{U}\boldsymbol{V}^*\right)\right] \boldsymbol{V} \\
\nabla_{\boldsymbol{V}} \mathcal{L}^\sigma(\nu, \boldsymbol{U}, \boldsymbol{V}) &= -\left[\varphi_{2\sigma^2} * \left(X_\natural \circ \boldsymbol{\tau}_{\nu - \nu_\natural} - \boldsymbol{U}\boldsymbol{V}^*\right)\right]^* \boldsymbol{U}.
\end{aligned}$$

It is convenient to simplify the adjoint operation in the second expression. First, note that $f \mapsto [\varphi_{2\sigma^2} * (X_\natural \circ \boldsymbol{\tau}_{\nu - \nu_\natural} - \boldsymbol{U}\boldsymbol{V}^*)][f]$ is a bounded operator on $L^2(\mathbb{R})$, because its $L^2 \to L^2$ operator norm is bounded by its Hilbert-Schmidt norm, which is finite:

$$\left\|\varphi_{2\sigma^2} * (X_\natural \circ \boldsymbol{\tau}_{\nu - \nu_\natural} - \boldsymbol{U}\boldsymbol{V}^*)\right\|_{\mathrm{HS}}^2 = \int_{\mathbb{R}^2} [\varphi_{2\sigma^2} * (X_\natural \circ \boldsymbol{\tau}_{\nu - \nu_\natural} - \boldsymbol{U}\boldsymbol{V}^*)](s,t)^2 \,\mathrm{d}s\,\mathrm{d}t < +\infty,$$

by Young's inequality for convolutions. This allows us to use Fubini's theorem freely in the sequel. For any $f, g \in L^2(\mathbb{R})$, we have

$$\begin{aligned}
\left\langle \left[\varphi_{2\sigma^2} * \left(X_\natural \circ \boldsymbol{\tau}_{\nu - \nu_\natural} - \boldsymbol{U}\boldsymbol{V}^*\right)\right][f], g \right\rangle_{L^2(\mathbb{R})} &= \int_{\mathbb{R}} g(s) \left(\int_{\mathbb{R}} \left[\varphi_{2\sigma^2} * (X_\natural \circ \boldsymbol{\tau}_{\nu - \nu_\natural} - \boldsymbol{U}\boldsymbol{V}^*)\right](s,t) f(t) \,\mathrm{d}t \right) \mathrm{d}s \\
&= \int_{\mathbb{R}} \int_{\mathbb{R}} \int_{\mathbb{R}^2} f(t) g(s) \varphi_{2\sigma^2}((s,t) - \boldsymbol{x}) (X_\natural \circ \boldsymbol{\tau}_{\nu - \nu_\natural} - \boldsymbol{U}\boldsymbol{V}^*)(\boldsymbol{x}) \,\mathrm{d}\boldsymbol{x}\,\mathrm{d}t\,\mathrm{d}s \\
&= \int_{\mathbb{R}} \int_{\mathbb{R}} \int_{\mathbb{R}^2} f(t) g(s) \varphi_{2\sigma^2}((t,s) - \boldsymbol{x}) (X_\natural \circ \boldsymbol{\tau}_{\nu - \nu_\natural} - \boldsymbol{U}\boldsymbol{V}^*) \left(\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \boldsymbol{x} \right) \mathrm{d}\boldsymbol{x}\,\mathrm{d}t\,\mathrm{d}s,
\end{aligned} \tag{F.60}$$

where the third equality uses a unitary change of variables $\boldsymbol{x} \mapsto \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \boldsymbol{x}$ in the convolution integral. Notice that

$$\begin{aligned}
(X_\natural \circ \boldsymbol{\tau}_{\nu - \nu_\natural} - \boldsymbol{U}\boldsymbol{V}^*)(s,t) &= X_\natural \circ \boldsymbol{\tau}_{\nu - \nu_\natural}(s,t) - \sum_{i=1}^k u_i(s) v_i(t) \\
&= X_\natural \left(\boldsymbol{R}_{\nu - \nu_\natural} \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} (t,s) \right) - \sum_{i=1}^k v_i(t) u_i(s).
\end{aligned}$$

Moreover, if $\boldsymbol{Q}$ is any orthogonal matrix with determinant -1, then for any ν one has $\det(\boldsymbol{R}_\nu \boldsymbol{Q}) = -1$; because the orthogonal matrices form a Lie group and every 2×2 orthogonal matrix with determinant -1 is symmetric, it follows

$$\boldsymbol{R}_\nu \boldsymbol{Q} = \boldsymbol{Q} \boldsymbol{R}_\nu^*. \tag{F.61}$$

In particular,

$$\boldsymbol{R}_{\nu - \nu_\natural} \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \boldsymbol{R}_{\nu - \nu_\natural}^*,$$

which, together with the fact that $X_\natural(s,t) = X_\natural(t,s)$, implies that

$$(X_\natural \circ \boldsymbol{\tau}_{\nu - \nu_\natural} - \boldsymbol{U}\boldsymbol{V}^*)(s,t) = (X_\natural \circ \boldsymbol{\tau}_{\nu_\natural - \nu} - \boldsymbol{V}\boldsymbol{U}^*)(t,s).$$

Applying this to (F.60) and unwinding the preceding steps implies immediately

$$\left\langle \left[\varphi_{2\sigma^2} * \left(X_\natural \circ \boldsymbol{\tau}_{\nu - \nu_\natural} - \boldsymbol{U}\boldsymbol{V}^*\right)\right][f], g \right\rangle_{L^2(\mathbb{R})} = \left\langle \left[\varphi_{2\sigma^2} * \left(X_\natural \circ \boldsymbol{\tau}_{\nu_\natural - \nu} - \boldsymbol{V}\boldsymbol{U}^*\right)\right][g], f \right\rangle_{L^2(\mathbb{R})},$$

which implies the claimed expression for the gradients with respect to $\boldsymbol{V}$. The gradient with respect to ν is a similar calculation with the chain rule, but involves some simplifications so we reproduce it here. From the chain rule, for any $\Delta\nu \in \mathbb{R}$ we have for the differential

$$\begin{aligned}
\mathrm{d}_\nu[\mathcal{L}^\sigma(\,\cdot\,, \boldsymbol{U}, \boldsymbol{V})](\Delta\nu) &= \left\langle \varphi_{\sigma^2} * \left(X_\natural \circ \boldsymbol{\tau}_{\nu-\nu_\natural} - \boldsymbol{U}\boldsymbol{V}^*\right), \frac{\partial}{\partial t}\bigg|_{t=0} [X_\natural \circ \boldsymbol{\tau}_{\nu+t\Delta\nu-\nu_\natural}] \right\rangle_{L^2(\mathbb{R}^2)} \\
&= \left\langle \varphi_{\sigma^2} * \left(X_\natural \circ \boldsymbol{\tau}_{\nu-\nu_\natural} - \boldsymbol{U}\boldsymbol{V}^*\right), \left\langle \nabla_{\boldsymbol{x}}[\varphi_{\sigma^2} * X_\natural] \circ \boldsymbol{\tau}_{\nu-\nu_\natural}, \dot{\boldsymbol{R}}_{\nu-\nu_\natural}(\,\cdot\,) \right\rangle_{\ell^2} \right\rangle_{L^2(\mathbb{R}^2)} \Delta\nu, \qquad \text{(F.62)}
\end{aligned}$$

where $\dot{\boldsymbol{R}}_\nu$ is the elementwise derivative of the expression in (F.2) for $\boldsymbol{R}_\nu$, which evaluates as

$$\begin{aligned}
\dot{\boldsymbol{R}}_\nu &= \begin{bmatrix} -\sin\nu & -\cos\nu \\ \cos\nu & -\sin\nu \end{bmatrix} \\
&= \begin{bmatrix} \cos\nu & -\sin\nu \\ \sin\nu & \cos\nu \end{bmatrix} \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} \\
&= \boldsymbol{R}_\nu \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}.
\end{aligned}$$

Note that the expression in the ℓ^2 inner product in (F.62) is a function of $\boldsymbol{x} \in \mathbb{R}^2$. In particular, the function

$$\boldsymbol{x} \mapsto \left\langle \boldsymbol{R}^*_{\nu-\nu_\natural} \nabla_{\boldsymbol{x}}[\varphi_{\sigma^2} * X_\natural] \circ \boldsymbol{\tau}_{\nu-\nu_\natural}(\boldsymbol{x}), \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} \boldsymbol{x} \right\rangle_{\ell^2}$$

gives the rotational component (tangential to the co-incident circle centered at the origin) of the rotated gradient vector field of $\varphi_{\sigma^2} * X_\natural$ at the point $\boldsymbol{x}$. This gives the expression

$$\nabla_\nu \mathcal{L}^\sigma(\nu, \boldsymbol{U}, \boldsymbol{V}) = \left\langle \varphi_{\sigma^2} * \left(X_\natural \circ \boldsymbol{\tau}_{\nu-\nu_\natural} - \boldsymbol{U}\boldsymbol{V}^*\right), \left\langle \boldsymbol{R}^*_{\nu-\nu_\natural} \nabla_{\boldsymbol{x}}[\varphi_{\sigma^2} * X_\natural] \circ \boldsymbol{\tau}_{\nu-\nu_\natural}, \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} (\,\cdot\,) \right\rangle_{\ell^2} \right\rangle_{L^2(\mathbb{R}^2)}.$$

Using the commutation relationship (F.31) and a unitary change of variables $\boldsymbol{x} \mapsto \boldsymbol{\tau}_{\nu_\natural-\nu}(\boldsymbol{x})$, the previous expression implies

$$\nabla_\nu \mathcal{L}^\sigma(\nu, \boldsymbol{U}, \boldsymbol{V}) = \left\langle \varphi_{\sigma^2} * \left(X_\natural - \boldsymbol{U}\boldsymbol{V}^* \circ \boldsymbol{\tau}_{\nu_\natural-\nu}\right), \left\langle \nabla_{\boldsymbol{x}}[\varphi_{\sigma^2} * X_\natural], \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} (\,\cdot\,) \right\rangle_{\ell^2} \right\rangle_{L^2(\mathbb{R}^2)}.$$

As in Lemma F.8, let $\mathcal{C}(\boldsymbol{x})$ denote the function of $\boldsymbol{x}$ encompassed by the ℓ^2 inner product. By Lemma F.8, we have that $\mathcal{C}(s,t) = -\mathcal{C}(t,s)$, whereas by (F.31) we have that $(\varphi_{\sigma^2} * X_\natural)(s,t) = (\varphi_{\sigma^2} * X_\natural)(t,s)$. It follows that these two functions are orthogonal over $L^2(\mathbb{R}^2)$, so that in particular

$$\nabla_\nu \mathcal{L}^\sigma(\nu, \boldsymbol{U}, \boldsymbol{V}) = -\left\langle \varphi_{\sigma^2} * \left(\boldsymbol{U}\boldsymbol{V}^* \circ \boldsymbol{\tau}_{\nu_\natural-\nu}\right), \left\langle \nabla_{\boldsymbol{x}}[\varphi_{\sigma^2} * X_\natural], \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} (\,\cdot\,) \right\rangle_{\ell^2} \right\rangle_{L^2(\mathbb{R}^2)},$$

as claimed. □

Lemma F.8. *Let*

$$\mathcal{C}(\boldsymbol{x}) = \left\langle \nabla_{\boldsymbol{x}}[\varphi_{\sigma^2} * X_\natural](\boldsymbol{x}), \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} \boldsymbol{x} \right\rangle_{\ell^2}$$

denote the rotational component of the gradient vector field of the smoothed template. Let $\boldsymbol{Q} \in \mathrm{O}(2)$ satisfy $\det(\boldsymbol{Q}) = -1$, and suppose that $\boldsymbol{Q}$ is a symmetry of the square template $X_\natural$: in particular

$$\boldsymbol{Q} \in \left\{ \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}, \begin{bmatrix} -1 & 0 \\ 0 & 1 \end{bmatrix}, \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, \begin{bmatrix} 0 & -1 \\ -1 & 0 \end{bmatrix} \right\} \qquad \text{(F.63)}$$

(this is the subgroup of D_4 consisting of symmetries of determinant -1). Then one has

$$\mathcal{C}(\boldsymbol{Q}\boldsymbol{x}) = -\mathcal{C}(\boldsymbol{x}).$$

Proof. With (F.2), we can write

$$\mathcal{C}(\boldsymbol{x}) = \left\langle \nabla_{\boldsymbol{x}}[\varphi_{\sigma^2} * X_\natural](\boldsymbol{x}), \boldsymbol{R}_{\pi/2}\boldsymbol{x} \right\rangle_{\ell^2}.$$

For $\boldsymbol{Q}$ as in (F.63) and using (F.61), we have

$$\begin{aligned}
\mathcal{C}(\boldsymbol{Q}\boldsymbol{x}) &= \left\langle \nabla_{\boldsymbol{x}}[\varphi_{\sigma^2} * X_\natural](\boldsymbol{Q}\boldsymbol{x}), \boldsymbol{Q}\boldsymbol{R}_{-\pi/2}\boldsymbol{x} \right\rangle_{\ell^2} \\
&= -\left\langle \nabla_{\boldsymbol{x}}[\varphi_{\sigma^2} * X_\natural](\boldsymbol{Q}\boldsymbol{x}), \boldsymbol{Q}\boldsymbol{R}_{\pi/2}\boldsymbol{x} \right\rangle_{\ell^2} \\
&= -\left\langle \boldsymbol{Q}\nabla_{\boldsymbol{x}}[\varphi_{\sigma^2} * X_\natural](\boldsymbol{Q}\boldsymbol{x}), \boldsymbol{R}_{\pi/2}\boldsymbol{x} \right\rangle_{\ell^2}.
\end{aligned} \tag{F.64}$$

The second line uses (F.2), and the third uses that every member of (F.63) is symmetric. By Young's inequality, we have

$$\nabla_{\boldsymbol{x}}[\varphi_{\sigma^2} * X_\natural] = \nabla_{\boldsymbol{x}}[\varphi_{\sigma^2}] * X_\natural,$$

and because φ_{σ^2} is invariant to all orthogonal matrices, its gradient is equivariant with respect to $\mathrm{O}(2)$, so in particular

$$\begin{aligned}
(\boldsymbol{Q}\nabla_{\boldsymbol{x}}[\varphi_{\sigma^2}] * X_\natural)(\boldsymbol{Q}\boldsymbol{x}) &= \int_{\mathbb{R}^2} X_\natural(\boldsymbol{x}')\boldsymbol{Q}\nabla_{\boldsymbol{x}}[\varphi_{\sigma^2}](\boldsymbol{Q}\boldsymbol{x} - \boldsymbol{x}')\,\mathrm{d}\boldsymbol{x}' \\
&= \int_{\mathbb{R}^2} X_\natural(\boldsymbol{x}')\nabla_{\boldsymbol{x}}[\varphi_{\sigma^2}](\boldsymbol{x} - \boldsymbol{Q}\boldsymbol{x}')\,\mathrm{d}\boldsymbol{x}' \\
&= \int_{\mathbb{R}^2} X_\natural(\boldsymbol{Q}\boldsymbol{x}')\nabla_{\boldsymbol{x}}[\varphi_{\sigma^2}](\boldsymbol{x} - \boldsymbol{x}')\,\mathrm{d}\boldsymbol{x}' \\
&= \int_{\mathbb{R}^2} X_\natural(\boldsymbol{x}')\nabla_{\boldsymbol{x}}[\varphi_{\sigma^2}](\boldsymbol{x} - \boldsymbol{x}')\,\mathrm{d}\boldsymbol{x}' \\
&= (\nabla_{\boldsymbol{x}}[\varphi_{\sigma^2}] * X_\natural)(\boldsymbol{x}).
\end{aligned}$$

Above, the third line uses an orthogonal change of variables in the convolution integral, and the fourth uses that $\boldsymbol{Q}$ is a symmetry of $X_\natural$. By (F.64), we have that $\mathcal{C}(\boldsymbol{Q}\boldsymbol{x}) = -\mathcal{C}(\boldsymbol{x})$.

□

Lemma F.9. *The following symmetry properties hold:*

1. *For any $u \in L^2(\mathbb{R})$, $\sigma > 0$, the objective $\nu \mapsto \mathcal{L}^\sigma(\nu, u)$ is $\pi/2$-periodic;*
2. *For any $u \in L^2(\mathbb{R})$, $\sigma > 0$, and $-\pi/4 \le \nu - \nu_\natural \le \pi/4$, one has $\mathcal{L}^\sigma(\nu - \nu_\natural, u) = \mathcal{L}^\sigma(\nu_\natural - \nu, u)$.*

Moreover, consider the alignment gradient at the nominal rough initial representation:

$$\nu \mapsto \nabla_\nu \mathcal{L}^\sigma(\nu, \bar{u}_{\text{rough}}).$$

where the nominal rough initial representation is defined as

$$\bar{u}_{\text{rough}}(s) = \frac{2}{\sqrt{\pi}} \mathbb{1}_{|s| \le 1} \cos(\pi s/2).$$

Suppose that $\alpha = \frac{1}{\sqrt{2}}$. Then for any ν,

$$|\nabla_\nu \mathcal{L}^\sigma(\nu, \bar{u}_{\text{rough}})| \le 256 |\sin(\nu - \nu_\natural)|.$$

If, in addition, $\sigma \le 10^{-3}$, then if ν is sufficiently far from maximizers, i.e., if

$$\left| \left(\nu - \nu_\natural + \frac{\pi}{4} \mod \frac{\pi}{2} \right) - \frac{\pi}{4} \right| \le \frac{\pi}{7}, \tag{F.65}$$

one has

$$\operatorname{sign}\left(\left(\nu - \nu_\natural + \frac{\pi}{4} \mod \frac{\pi}{2} \right) - \frac{\pi}{4} \right) \cdot \nabla_\nu \mathcal{L}^\sigma(\nu, \bar{u}_{\text{rough}}) \ge c_0 \sin\left(\left| \left(\nu - \nu_\natural + \frac{\pi}{4} \mod \frac{\pi}{2} \right) - \frac{\pi}{4} \right| \right)$$

for an absolute constant $c_0 > 0$. In particular, the gradient is nonnegative when $\nu - \nu_\natural \mod \pi/2 \le \pi/7$, and nonpositive when $\nu - \nu_\natural \mod \pi/2 \ge \pi/2 - \pi/7$.

Proof. The proof exploits heavily the D_4 symmetries of the square template $X_\natural$ (c.f. Lemma F.8) and of the initialization $\bar{u}_{\text{rough}}\bar{u}^*_{\text{rough}}$. Before proceeding with the analysis of the gradient, we go through some simplifying reductions based on symmetry. First, by the definition of the loss in (F.34), it suffices to analyze the case where $\nu_\natural = 0$, and perform the substitution $\nu \mapsto \nu - \nu_\natural$ in all results obtained. Next, notice that because $\boldsymbol{R}_{\nu+\pi/2} = \boldsymbol{R}_{\pi/2}\boldsymbol{R}_\nu$ for any ν (following the notation of (F.2)), one has $X_\natural(\boldsymbol{R}_{\nu+\pi/2}\boldsymbol{x}) = X_\natural(\boldsymbol{R}_\nu \boldsymbol{x})$ for any $\boldsymbol{x}$ by symmetry, which implies (c.f. (F.34)) that $\mathcal{L}^\sigma(\nu+\pi/2, u) = \mathcal{L}^\sigma(\nu, u)$ for any ν, u. Moreover, by (F.61), one has $\boldsymbol{R}_{-\nu} = \boldsymbol{Q}\boldsymbol{R}_\nu\boldsymbol{Q}$, where

$$\boldsymbol{Q} = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}.$$

We have that $\boldsymbol{Q}$ is an orthogonal matrix with determinant -1; writing $\boldsymbol{\tau}_{\boldsymbol{Q}} : \mathbb{R}^2 \to \mathbb{R}^2$ for its induced transformation, we have again by symmetry that $X_\natural \circ \boldsymbol{\tau}_{\pi/2-\nu}(\boldsymbol{x}) = X_\natural \circ \boldsymbol{\tau}_\nu \circ \boldsymbol{\tau}_{\boldsymbol{Q}}$. Applying then (F.31) (notice that the calculation does not use the fact that $\det(\boldsymbol{R}_\nu) = 1$, and in fact any orthogonal matrix yields the same conclusion) together with a unitary change of coordinates, we obtain that

$$\mathcal{L}^\sigma(\pi/2 - \nu, \bar{u}_{\text{rough}}) = \frac{1}{2}\left\|\varphi_{\sigma^2} * \left(X_\natural \circ \boldsymbol{\tau}_\nu - \bar{u}_{\text{rough}}\bar{u}^*_{\text{rough}} \circ \boldsymbol{\tau}_{\boldsymbol{Q}}\right)\right\|^2_{L^2},$$

since $\boldsymbol{Q}^* = \boldsymbol{Q}$. But since $\boldsymbol{\tau}_{\boldsymbol{Q}}(s,t) = (t,s)$, it follows from symmetry that $\mathcal{L}^\sigma(\pi/2-\nu, \bar{u}_{\text{rough}}) = \mathcal{L}^\sigma(\nu, \bar{u}_{\text{rough}})$. We have thus shown that $\nu \mapsto \mathcal{L}^\sigma(\nu, \bar{u}_{\text{rough}})$ is

1. $\pi/2$-periodic;
2. on $[0, \pi/2]$, symmetric about $\pi/4$.

It therefore suffices to assume that $0 \le \nu \le \pi/4$ in the sequel, since conclusions on this interval can be translated to all $\nu \in \mathbb{R}$ as stated in the statement of the result by these symmetry properties.

We proceed to estimate the gradient

$$\nabla_\nu \mathcal{L}^\sigma(\nu, \bar{u}_{\text{rough}}) = -\left\langle \underbrace{\varphi_{\sigma^2} * \left(\bar{u}_{\text{rough}}\bar{u}^*_{\text{rough}} \circ \boldsymbol{\tau}_{-\nu}\right)}_{\mathcal{R}}, \mathcal{C} \right\rangle_{L^2(\mathbb{R}^2)}$$

under the preceding assumptions, where $\mathcal{C}$ is defined as in Lemma F.8. First, we reduce the L^2 integral in the expression for a gradient into a difference of integrals over a 'fundamental domain' depending on the $\mathcal{D}_4$ symmetries of $X_\natural$ and the initialization; this expression will be useful for upper and lower bounds. Then, we will establish the lower bound, which is more technical, before concluding with the upper bound. For $(\boldsymbol{\kappa}, \pi) \in \{-1,1\}^2 \times \mathrm{P}(2)$, where $\mathrm{P}(2)$ is the set of permutations on 2 elements, we consider the "wedge" domains

$$C_{\boldsymbol{\kappa},\pi} = \left\{\boldsymbol{x} = (s,t) \in \mathbb{R}^2 \;\middle|\; \kappa_1 s \ge 0, \kappa_2 t \ge 0, \pi_1(s, \kappa_1\kappa_2 t) \ge \pi_2(s, \kappa_1\kappa_2 t)\right\}.$$

Intuitively, in the third constraint, π governs the "direction" of the inequality, and $\kappa_1\kappa_2$ selects the proper subspace to reflect about. One notes that $\cup_{(\boldsymbol{\kappa},\pi)\in\{-1,1\}^2\times\mathrm{P}(2)} C_{\boldsymbol{\kappa},\pi} = \mathbb{R}^2$, and if $(\boldsymbol{\kappa},\pi) \neq (\boldsymbol{\kappa}', \pi')$ then $C_{\boldsymbol{\kappa},\pi} \cap C_{\boldsymbol{\kappa}',\pi'}$ has zero Lebesgue measure. Because $\mathcal{C}$ and $\mathcal{R}$ are smooth functions, it follows

$$\begin{aligned}
\nabla_\nu \mathcal{L}^\sigma(\nu, \bar{u}_{\text{rough}}) &= -\left\langle \mathcal{R}\left(\sum_{(\boldsymbol{\kappa},\pi)\in\{-1,1\}^2\times\mathrm{P}(2)} \mathbb{1}_{C_{\boldsymbol{\kappa},\pi}}\right), \mathcal{C}\right\rangle_{L^2} \\
&= \sum_{(\boldsymbol{\kappa},\pi)\in\{-1,1\}^2\times\mathrm{P}(2)} -\left\langle \mathcal{R}\mathbb{1}_{C_{\boldsymbol{\kappa},\pi}}, \mathcal{C}\right\rangle_{L^2}.
\end{aligned}$$

Recall, following (F.31), that we can freely interchange the order of gaussian smoothing and rotation in the expression for $\mathcal{L}^\sigma(\nu, \bar{u}_{\text{rough}})$. Since $\bar{u}_{\text{rough}}(s) = \bar{u}_{\text{rough}}(-s)$ for any $s \in \mathbb{R}$, by the argument above applied to $X_\natural$ we have that for any ν

$$\begin{aligned}
\mathcal{R} \circ \boldsymbol{\tau}_{\pm\pi/2} &= \left(\varphi_{\sigma^2} * \left(\bar{u}_{\text{rough}}\bar{u}^*_{\text{rough}} \circ \boldsymbol{\tau}_{-\nu}\right)\right) \circ \boldsymbol{\tau}_{\pm\pi/2} \\
&= \varphi_{\sigma^2} * \left(\bar{u}_{\text{rough}}\bar{u}^*_{\text{rough}} \circ \boldsymbol{\tau}_{\pm\pi/2} \circ \boldsymbol{\tau}_{-\nu}\right) \\
&= \varphi_{\sigma^2} * \left(\bar{u}_{\text{rough}}\bar{u}^*_{\text{rough}} \circ \boldsymbol{\tau}_{-\nu}\right) \\
&= \mathcal{R}.
\end{aligned}$$

Meanwhile, we note that

$$\boldsymbol{R}_{\pi/2} = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}; \quad \boldsymbol{R}_{-\pi/2} = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} -1 & 0 \\ 0 & 1 \end{bmatrix},$$

so by Lemma F.8, $\mathcal{C} \circ \boldsymbol{\tau}_{\pm\pi/2} = \mathcal{C}$. Thus, changing coordinates in the L^2 integral, we get

$$\nabla_\nu \mathcal{L}^\sigma(\nu, \bar{u}_{\text{rough}}) = -4\left\langle \mathcal{R}\left(\mathbb{1}_{C_{\{1,-1\},\text{Id}}} + \mathbb{1}_{C_{\{1,1\},\text{Id}}}\right), \mathcal{C}\right\rangle_{L^2},$$

where we recall

$$C_{\{1,-1\},\text{Id}} = \left\{\boldsymbol{x} = (s,t) \in \mathbb{R}^2 \;\middle|\; s \geq 0, t \leq 0, s \geq -t\right\}, \quad C_{\{1,1\},\text{Id}} = \left\{\boldsymbol{x} = (s,t) \in \mathbb{R}^2 \;\middle|\; s \geq 0, t \geq 0, s \geq t\right\}. \tag{F.66}$$

By another change of coordinates and Lemma F.8, we then have in addition

$$\nabla_\nu \mathcal{L}^\sigma(\nu, \bar{u}_{\text{rough}}) \geq -4 \int_{\{0 \leq t \leq s\}} \left(\mathcal{R}(s,t) - \mathcal{R}(s,-t)\right) \mathcal{C}(s,t)\, \mathrm{d}s\, \mathrm{d}t \tag{F.67}$$

Next, we control $\mathcal{R}(s,t) - \mathcal{R}(s,-t)$ using the geometry of the rough factorization $\bar{u}_{\text{rough}}\bar{u}^*_{\text{rough}}$; we will then conclude the bound from (F.67). First, by linearity and (F.31), we have

$$\begin{aligned} \mathcal{R}(s,t) - \mathcal{R}(s,-t) &= \varphi_{\sigma^2} * \left(\bar{u}_{\text{rough}}\bar{u}^*_{\text{rough}} \circ \boldsymbol{\tau}_{-\nu}\right)(s,t) - \varphi_{\sigma^2} * \left(\bar{u}_{\text{rough}}\bar{u}^*_{\text{rough}} \circ \boldsymbol{\tau}_{-\nu}\right)(s,-t) \\ &= \varphi_{\sigma^2} * \left(\bar{u}_{\text{rough}}\bar{u}^*_{\text{rough}} \circ \boldsymbol{\tau}_{-\nu} - \bar{u}_{\text{rough}}\bar{u}^*_{\text{rough}} \circ \boldsymbol{\tau}_{-\nu} \circ \boldsymbol{\tau}_{\boldsymbol{Q}}\right)(s,t), \end{aligned} \tag{F.68}$$

where $\boldsymbol{Q}$ is the matrix representation of the orthogonal transformation $(s,t) \mapsto (s,-t)$. Gaussian smoothing is nonnegativity-preserving, so developing a lower bound on the difference $\mathcal{R}(s,t) - \mathcal{R}(s,-t)$ can be done by developing a lower bound on the parenthesized term above. By (F.61) and symmetry, we have

$$\bar{u}_{\text{rough}}\bar{u}^*_{\text{rough}} \circ \boldsymbol{\tau}_{-\nu} \circ \boldsymbol{\tau}_{\boldsymbol{Q}} = \bar{u}_{\text{rough}}\bar{u}^*_{\text{rough}} \circ \boldsymbol{\tau}_{\nu}. \tag{F.69}$$

Applying the first conclusion in Lemma F.12 together with Lemma F.11, it follows that

$$-(\mathcal{R}(s,t) - \mathcal{R}(s,-t))\mathcal{C}(s,t) \geq 0$$

for every $0 \leq t \leq s$.This means that we can obtain a lower bound for the RHS of (F.67) by integrating over a subset of the domain $\{0 \leq t \leq s\}$. When σ is small, the field $\mathcal{C}$ concentrates around the boundary of the square template $X_\natural$; we will therefore obtain a lower bound for the gradient by integrating in a small strip around this region. To this end, the second conclusion in Lemma F.12 gives the following quantitative bound, valid for $0 \leq t \leq s \leq 1$ and all $0 \leq \nu \leq \pi/7$:

$$\bar{u}_{\text{rough}}\bar{u}^*_{\text{rough}} \circ \boldsymbol{\tau}_{\nu}(\boldsymbol{x}) - \bar{u}_{\text{rough}}\bar{u}^*_{\text{rough}} \circ \boldsymbol{\tau}_{-\nu}(\boldsymbol{x}) \geq \frac{7 \sin \nu}{1000} \mathbb{1}_{-0.137 \leq t - \frac{1}{\sqrt{2}} \leq -0.127} \mathbb{1}_{-0.001 \leq s - \frac{1}{\sqrt{2}} \leq 0.001}.$$

Since we are considering a regime with σ small, it is now reasonable to simplify this estimate further by worst-casing the smoothing that connects it to $\mathcal{R}$. Because this indicator is a box in the (s,t) plane, its smoothed version is a product of smoothed indicators for compact connected intervals in $\mathbb{R}$. If $I = [-a,a]$ is such an interval (because convolution commutes with translations, it will be sufficient to consider such a centered interval), we have (see the derivative calculations at the start of the proof of Lemma F.13) that $\varphi_{\sigma^2} * \mathbb{1}_I(x)$ is decreasing (resp. increasing) for $x \geq 0$ (resp. $x \leq 0$). Hence the minimum value taken by $\varphi_{\sigma^2} * \mathbb{1}_I$ among those $x \in I$ is attained at $x \in \{\pm a\}$, where

$$\begin{aligned} \mathbb{1}_I * \varphi_{\sigma^2}(a) = \int_{-a}^{a} \varphi_{\sigma^2}(x-a)\, \mathrm{d}x &= \int_0^{2a} \varphi_{\sigma^2}(x)\, \mathrm{d}x \\ &= \frac{1}{2} - \int_{2a}^{\infty} \varphi_{\sigma^2}(x)\, \mathrm{d}x \\ &\geq \frac{1}{2} - \frac{\sigma}{2a\sqrt{2\pi}} e^{-2a^2/\sigma^2}, \end{aligned}$$

by the standard estimate for the gaussian tail integral. Thus, as soon as $\sigma \le 2a$, one has

$$\mathbb{1}_I * \varphi_{\sigma^2}(a) \ge \frac{1}{4}, \tag{F.70}$$

which shows that $\mathbb{1}_I * \varphi_{\sigma^2} \ge \frac{1}{4}\mathbb{1}_I$ if $\sigma \le 2a$. Applying this to our lower bound, it follows that if $\sigma \le \frac{1}{500}$, we have

$$-\left(\mathcal{R}(s,t) - \mathcal{R}(s,-t)\right) \ge \frac{7\sin\nu}{16000}\mathbb{1}_{-0.137\le t-\frac{1}{\sqrt{2}}\le -0.127}\mathbb{1}_{-0.001\le s-\frac{1}{\sqrt{2}}\le 0.001}.$$

Plugging this bound into (F.67) gives

$$\nabla_\nu \mathcal{L}^\sigma(\nu, \bar{u}_{\text{rough}}) \ge \frac{7\sin\nu}{4000} \iint_{\substack{-0.001\le s-1/\sqrt{2}\le 0.001,\\ -0.137\le t-1/\sqrt{2}\le -0.127}} \mathcal{C}(s,t)\,\mathrm{d}s\,\mathrm{d}t. \tag{F.71}$$

The remainder of the proof is a relatively tedious calculation over this domain of integration. We make use of the expressions for $\mathcal{C}$ derived in Lemma F.11:

$$\begin{aligned}
\mathcal{C}(s,t) &= sf(s)f'(t) - tf(t)f'(s), \quad \text{where}\\
f(x) &= \frac{1}{\sqrt{2\pi\sigma^2}}\int_{-1/\sqrt{2}}^{1/\sqrt{2}} e^{-\frac{(x-x')^2}{2\sigma^2}}\,\mathrm{d}x', \quad x\in\mathbb{R};\\
f'(x) &= \varphi_{\sigma^2}(x+\alpha) - \varphi_{\sigma^2}(x-\alpha).
\end{aligned}$$

Moreover, we recall that $f'(x) \ge 0$ if $x \ge 0$ and $f'(x) \le 0$ if $x \le 0$, as shown in the proof of Lemma F.11. We have $|sf(s)| \le 1$ when $s \le 1$ by Young's convolution inequality, and

$$\begin{aligned}
f'(t) &\ge -\varphi_{\sigma^2}(t - 1/\sqrt{2})\\
&\ge -\varphi_{\sigma^2}(-0.127)\\
&\ge -\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{1}{128\sigma^2}}
\end{aligned}$$

for t in the region of integration. Similarly, for t in the region of integration

$$\begin{aligned}
tf(t) &\ge \left(\frac{1}{\sqrt{2}} - 0.137\right) f(1/\sqrt{2} - 0.127)\\
&\ge \frac{1}{4}\left(\frac{1}{\sqrt{2}} - 0.137\right),
\end{aligned}$$

where the last line uses (F.70). Finally, we have for s in the domain of integration

$$\begin{aligned}
f'(s) &= \varphi_{\sigma^2}(s - 1/\sqrt{2}) - \varphi_{\sigma^2}(s + 1/\sqrt{2})\\
&\ge \varphi_{\sigma^2}(s - 1/\sqrt{2}) - \varphi_{\sigma^2}(\sqrt{2} - 0.001)\\
&\ge \varphi_{\sigma^2}(s - 1/\sqrt{2}) - \frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{1}{2\sigma^2}}.
\end{aligned}$$

Combining these, we have the lower bound (valid on the domain of integration of our gradient lower bound)

$$\begin{aligned}
\mathcal{C}(s,t) &\ge \frac{1}{4}\left(\frac{1}{\sqrt{2}} - 0.137\right)\left(\varphi_{\sigma^2}(s - 1/\sqrt{2}) - \frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{1}{2\sigma^2}}\right) - \frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{1}{128\sigma^2}}\\
&\ge \frac{1}{8}\varphi_{\sigma^2}(s - 1/\sqrt{2}) - \frac{1}{\sigma\sqrt{\pi}}e^{-\frac{1}{128\sigma^2}}.
\end{aligned}$$

Integrating the first term in this lower bound over the s region gives, by a change of coordinates,

$$\begin{aligned}\int_{-0.001+1/\sqrt{2}}^{0.001+1/\sqrt{2}} \varphi_{\sigma^2}(s-1/\sqrt{2})\,\mathrm{d}s &= \int_{-0.001}^{0.001} \varphi_{\sigma^2}(s)\,\mathrm{d}s \\ &= 1 - 2\int_0^{0.001} \varphi_{\sigma^2}(s)\,\mathrm{d}s \\ &\geq 1 - 2\frac{\sigma}{10^{-3}\sqrt{2\pi}} e^{-10^{-6}/2\sigma^2},\end{aligned}$$

using also the gaussian tail estimate we applied above. Thus, as soon as $\sigma \leq 10^{-3}$, we have

$$\int_{-0.001+1/\sqrt{2}}^{0.001+1/\sqrt{2}} \varphi_{\sigma^2}(s-1/\sqrt{2})\,\mathrm{d}s \geq \frac{1}{2},$$

and under this constraint on σ, we have moreover from our previous lower bound

$$\int_{-0.001+1/\sqrt{2}}^{0.001+1/\sqrt{2}} \mathcal{C}(s,t)\,\mathrm{d}s \geq \frac{1}{20}.$$

Integrating over the region of t adds only an additional constant multiple, since this expression does not depend on t. Consequently, these calculations together with (F.71) imply the claimed lower bound.

We can obtain the claimed upper bound in a similar way. Since the lower bound we have just established characterizes the sign of the gradient at all points where $0 \leq \nu \leq \pi/7$ (and similarly for negative ν, by symmetry of the objective), it suffices to simply control the magnitude of the gradient. By (F.67), (F.69) and (F.68), and L^1-L^∞ control, we have

$$\begin{aligned}|\nabla_\nu \mathcal{L}^\sigma(\nu, \bar{u}_{\text{rough}})| &\leq 4\left(\sup_{0\leq t\leq s} \left|\varphi_{\sigma^2} * \left(\bar{u}_{\text{rough}}\bar{u}^*_{\text{rough}} \circ \boldsymbol{\tau}_{-\nu} - \bar{u}_{\text{rough}}\bar{u}^*_{\text{rough}} \circ \boldsymbol{\tau}_{\nu}\right)(s,t)\right|\right) \int_{\{s\geq 0, t\geq 0, s\geq t\}} |\mathcal{C}(s,t)|\,\mathrm{d}s\,\mathrm{d}t \\ &\leq 4\left(\sup_{0\leq t\leq s} \left|\bar{u}_{\text{rough}}\bar{u}^*_{\text{rough}} \circ \boldsymbol{\tau}_{-\nu} - \bar{u}_{\text{rough}}\bar{u}^*_{\text{rough}} \circ \boldsymbol{\tau}_{\nu}\right|(s,t)\right) \int_{\{0\leq t\leq s\}} |\mathcal{C}(s,t)|\,\mathrm{d}s\,\mathrm{d}t,\end{aligned}$$

where we use Young's convolution inequality in the second line. By the expressions given above and the triangle inequality, we have

$$\begin{aligned}\int_{\{0\leq t\leq s\}} |\mathcal{C}(s,t)|\,\mathrm{d}s\,\mathrm{d}t &\leq \int_{\{0\leq t\leq s\}} (|sf(s)f'(t)| + |tf(t)f'(s)|)\,\mathrm{d}s\,\mathrm{d}t \\ &\leq 2\int_{\mathbb{R}^2} (|sf(s)f'(t)|)\,\mathrm{d}s\,\mathrm{d}t \\ &= 2\left(\int_{\mathbb{R}} |sf(s)|\,\mathrm{d}s\right)\left(\int_{\mathbb{R}} |f'(t)|\,\mathrm{d}t\right),\end{aligned}$$

where the third line uses Fubini's theorem. Because f' is a difference of two gaussians, the integral of its magnitude is no larger than 2. Meanwhile, we have by Fubini's theorem

$$\begin{aligned}\int_{\mathbb{R}} |sf(s)|\,\mathrm{d}s &= \int_{-1/\sqrt{2}}^{1/\sqrt{2}} \int_{\mathbb{R}} |s|\varphi_{\sigma^2}(s-x)\,\mathrm{d}s\,\mathrm{d}x \\ &= \int_{-1/\sqrt{2}}^{1/\sqrt{2}} \int_{\mathbb{R}} |s+x|\varphi_{\sigma^2}(s)\,\mathrm{d}s\,\mathrm{d}x \\ &\leq \int_{-1/\sqrt{2}}^{1/\sqrt{2}} (|x| + \sqrt{2/\pi})\,\mathrm{d}x \\ &= \frac{1}{2} + \frac{2}{\sqrt{\pi}}.\end{aligned}$$

Thus

$$\int_{\{0\le t\le s\}} |\mathcal{C}(s,t)|\,\mathrm{d}s\,\mathrm{d}t \le 8.$$

Meanwhile, by definition (see Lemma F.9), $\bar{u}_{\text{rough}}$ is a $\sqrt{\pi}$-Lipschitz function of its argument, and is bounded by $2/\sqrt{\pi}$; this means that $\bar{u}_{\text{rough}}\bar{u}^*_{\text{rough}} : \mathbb{R}^2 \to \mathbb{R}$ is a $2\sqrt{2}$-Lipschitz function of its argument with respect to the ℓ^2 metric on $\mathbb{R}^2$. Recalling moreover that $\bar{u}_{\text{rough}}$ is compactly supported on $[-1,1]$, it follows that for any $\boldsymbol{x} = (s,t)$ at which $\bar{u}_{\text{rough}}\bar{u}^*_{\text{rough}} \circ \boldsymbol{\tau}_{\pm\nu}$ is nonzero, we have

$$\begin{aligned}\left|\bar{u}_{\text{rough}}\bar{u}^*_{\text{rough}} \circ \boldsymbol{\tau}_{-\nu}(\boldsymbol{x}) - \bar{u}_{\text{rough}}\bar{u}^*_{\text{rough}} \circ \boldsymbol{\tau}_{\nu}(\boldsymbol{x})\right| &\le 2\sqrt{2}\|\boldsymbol{\tau}_{-\nu}(\boldsymbol{x}) - \boldsymbol{\tau}_{\nu}(\boldsymbol{x})\|_2 \\ &\le 2\sqrt{2}\|\boldsymbol{R}_{-\nu} - \boldsymbol{R}_{\nu}\|\|\boldsymbol{x}\|_2 \\ &\le 4\|\boldsymbol{R}_{-\nu} - \boldsymbol{R}_{\nu}\|.\end{aligned}$$

In the last two lines, we simply pass to the operator norm of the difference of rotation matrices and then use that $\|\boldsymbol{x}\|_\infty \le 1$. In dimension two, we have the representation

$$\boldsymbol{R}_\nu = (\cos\nu)\boldsymbol{I} + (\sin\nu)\begin{bmatrix}0 & -1\\ 1 & 0\end{bmatrix},$$

and each of these matrices (without the prefactors) is orthogonal, hence has unit operator norm. Thus, the triangle inequality gives

$$\|\boldsymbol{R}_{-\nu} - \boldsymbol{R}_{\nu}\| \le 2|\sin\nu|.$$

since sin and cos are both 1-Lipschitz. Combining, this shows

$$|\nabla_\nu \mathcal{L}^\sigma(\nu, \bar{u}_{\text{rough}})| \le 256\sin\nu.$$

□

Lemma F.10. *Consider the roughly-localized alignment iteration (F.41). Suppose*

$$\|u_{\text{rough}} - \bar{u}_{\text{rough}}\|_{L^2(\mathbb{R})} \le \frac{2}{\sqrt{\pi}},$$

where the nominal rough initial representation $\bar{u}_{\text{rough}}$ is defined as in Lemma F.9. For any $\sigma \le \frac{1}{100}$ and $\alpha = \frac{1}{\sqrt{2}}$, the following holds:

1. *The functions $\nabla_\nu \mathcal{L}^\sigma(\,\cdot\,, u_{\text{rough}})$ and $\nabla_\nu \mathcal{L}^\sigma(\,\cdot\,, \bar{u}_{\text{rough}})$ satisfy*

$$\begin{aligned}\|\nabla_\nu \mathcal{L}^\sigma(\,\cdot\,, \bar{u}_{\text{rough}})\|_{\text{Lip}} &\le \frac{4}{\pi\sigma^2}, \\ \|\nabla_\nu \mathcal{L}^\sigma(\,\cdot\,, u_{\text{rough}})\|_{\text{Lip}} &\le \frac{4}{\sigma^2}\left(\frac{1}{\pi} + \frac{2\|u_{\text{rough}} - \bar{u}_{\text{rough}}\|_{L^2(\mathbb{R})}}{\sqrt{\pi}}\right),\end{aligned}$$

   where $\|\cdot\|_{\text{Lip}}$ denotes the Lipschitz seminorm;

2. *We have the gradient estimate*

$$|\nabla_\nu \mathcal{L}^\sigma(\,\cdot\,, u_{\text{rough}}) - \nabla_\nu \mathcal{L}^\sigma(\,\cdot\,, \bar{u}_{\text{rough}})| \le \frac{8}{\sqrt{\pi}\sigma^2}\|\bar{u}_{\text{rough}} - u_{\text{rough}}\|_{L^2(\mathbb{R})},$$

   as well as the squared-gradient estimate

$$\left|(\nabla_\nu \mathcal{L}^\sigma(\,\cdot\,, u_{\text{rough}}))^2 - (\nabla_\nu \mathcal{L}^\sigma(\,\cdot\,, \bar{u}_{\text{rough}}))^2\right| \le \frac{512}{\pi^{3/2}\sigma^4}\|\bar{u}_{\text{rough}} - u_{\text{rough}}\|_{L^2(\mathbb{R})}.$$

Proof. We will prove the first assertion first. We recall from Lemma F.7 that for any $\nu \in \mathbb{R}$, $u \in L^2(\mathbb{R})$,

$$\nabla_\nu \mathcal{L}^\sigma(\nu, u) = -\left\langle \varphi_{\sigma^2} * \left(uu^* \circ \boldsymbol{\tau}_{\nu_\natural - \nu}\right), \left\langle \nabla_{\boldsymbol{x}}[\varphi_{\sigma^2} * X_\natural], \boldsymbol{\tau}_{\pi/2} \right\rangle_{\ell^2} \right\rangle_{L^2(\mathbb{R}^2)},$$

and using (F.31) and the fact that $f \mapsto f \circ \boldsymbol{\tau}_\nu$ is a unitary transformation of $L^2(\mathbb{R}^2)$, we thus have

$$\nabla_\nu \mathcal{L}^\sigma(\nu, u) = -\left\langle \varphi_{\sigma^2} * uu^*, \mathcal{C}^{u_\natural} \circ \boldsymbol{\tau}_{\nu - \nu_\natural} \right\rangle_{L^2(\mathbb{R}^2)},$$

using notation from Lemmas F.8 and F.13 as

$$\mathcal{C}^u(\boldsymbol{x}) = \left\langle \nabla_{\boldsymbol{x}}[\varphi_{\sigma^2} * uu^*](\boldsymbol{x}), \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} \boldsymbol{x} \right\rangle_{\ell^2}.$$

It then follows that

$$\begin{aligned}
\nabla_\nu^2 \mathcal{L}^\sigma(\nu, u) &= -\left\langle \varphi_{\sigma^2} * uu^*, \left\langle \boldsymbol{R}^*_{\nu - \nu_\natural} \nabla_{\boldsymbol{x}} \mathcal{C}^{u_\natural} \circ \boldsymbol{\tau}_{\nu - \nu_\natural}, \boldsymbol{\tau}_{\pi/2} \right\rangle_{\ell^2} \right\rangle_{L^2(\mathbb{R}^2)} \\
&= -\left\langle \varphi_{\sigma^2} * \left(uu^* \circ \boldsymbol{\tau}_{\nu_\natural - \nu}\right), \left\langle \nabla_{\boldsymbol{x}} \mathcal{C}^{u_\natural}, \boldsymbol{\tau}_{\pi/2} \right\rangle_{\ell^2} \right\rangle_{L^2(\mathbb{R}^2)},
\end{aligned}$$

calculating as in the proof of Lemma F.7 for the first derivative. We can estimate the RHS with the Schwarz inequality; by Young's convolution inequality and the fact that $f \mapsto f \circ \boldsymbol{\tau}_{-\nu}$ is a unitary transformation of $L^2(\mathbb{R}^2)$, we obtain

$$\begin{aligned}
\left\| \varphi_{\sigma^2} * \left(uu^* \circ \boldsymbol{\tau}_{\nu_\natural - \nu}\right) \right\|_{L^2(\mathbb{R}^2)} &\le \left\| \varphi_{\sigma^2} \right\|_{L^1(\mathbb{R}^2)} \left\| uu^* \circ \boldsymbol{\tau}_{\nu_\natural - \nu} \right\|_{L^2(\mathbb{R}^2)} \\
&\le \|uu^*\|_{L^2(\mathbb{R}^2)} \\
&= \|u\|^2_{L^2(\mathbb{R})}.
\end{aligned}$$

Meanwhile, by the second estimate in Lemma F.14, we have

$$\left\| \left\langle \nabla_{\boldsymbol{x}} \mathcal{C}^{u_\natural}, \boldsymbol{\tau}_{\pi/2} \right\rangle_{\ell^2} \right\|_{L^2(\mathbb{R}^2)} \le \left(\frac{3}{\pi} + \frac{55}{\pi\sigma^2} + \frac{4}{5\pi\sigma^4} \right)^{1/2} \le \frac{1}{\sigma^2},$$

where the worst-casing uses that $\sigma^2 \le \frac{1}{100}$. Thus, we have

$$\left| \nabla_\nu^2 \mathcal{L}^\sigma(\nu, u) \right| \le \frac{\|u\|^2_{L^2(\mathbb{R})}}{\sigma^2}.$$

For real numbers x, y, we have

$$|x^2 - y^2| \le 2 \max\{|x|, |y|\} |x - y|,$$

so by the triangle inequality,

$$\begin{aligned}
\left| \|u_{\text{rough}}\|^2_{L^2(\mathbb{R})} - \|\bar{u}_{\text{rough}}\|^2_{L^2(\mathbb{R})} \right| &\le 2 \max\left\{ \|u_{\text{rough}}\|_{L^2(\mathbb{R})}, \|\bar{u}_{\text{rough}}\|_{L^2(\mathbb{R})} \right\} \|u_{\text{rough}} - \bar{u}_{\text{rough}}\|_{L^2(\mathbb{R})} \\
&\le 2 \left(\|\bar{u}_{\text{rough}}\|_{L^2(\mathbb{R})} + \|u_{\text{rough}} - \bar{u}_{\text{rough}}\|_{L^2(\mathbb{R})} \right) \|u_{\text{rough}} - \bar{u}_{\text{rough}}\|_{L^2(\mathbb{R})} \\
&\le 4 \|\bar{u}_{\text{rough}}\|_{L^2(\mathbb{R})} \|u_{\text{rough}} - \bar{u}_{\text{rough}}\|_{L^2(\mathbb{R})},
\end{aligned}$$

where the last line requires that $\|u_{\text{rough}} - \bar{u}_{\text{rough}}\|_{L^2(\mathbb{R})} \le \|\bar{u}_{\text{rough}}\|_{L^2(\mathbb{R})}$. Since $\|\bar{u}_{\text{rough}}\|_{L^2(\mathbb{R})} = 2/\sqrt{\pi}$ by Lemma F.1, this, combined with our previously-derived bound, is equivalent to the assertion.

For the second assertion, we have by the above inequality

$$\begin{aligned}
\left| \left(\nabla_\nu \mathcal{L}^\sigma(\,\cdot\,, u_{\text{rough}})\right)^2 - \left(\nabla_\nu \mathcal{L}^\sigma(\,\cdot\,, \bar{u}_{\text{rough}})\right)^2 \right| \le& 2 \max\{ |\nabla_\nu \mathcal{L}^\sigma(\,\cdot\,, u_{\text{rough}})|, |\nabla_\nu \mathcal{L}^\sigma(\,\cdot\,, \bar{u}_{\text{rough}})|, \} \\
&\times |\nabla_\nu \mathcal{L}^\sigma(\,\cdot\,, u_{\text{rough}}) - \nabla_\nu \mathcal{L}^\sigma(\,\cdot\,, \bar{u}_{\text{rough}})|,
\end{aligned}$$

so we can bound the sizes of the two factors as well as their absolute difference. Using the expression given in the proof of the previous assertion for the gradient, we note that for any $\nu \in \mathbb{R}$

$$\nabla_\nu \mathcal{L}^\sigma(\nu, 0) = 0,$$

so bounding the sizes of the two factors is accomplished by a bound on their difference. Now, by linearity and the Schwarz inequality, we have for any any $u, v \in L^2(\mathbb{R})$ and any $\nu \in \mathbb{R}$

$$\begin{aligned}
|\nabla_\nu \mathcal{L}^\sigma(\nu, u) - \nabla_\nu \mathcal{L}^\sigma(\nu, v)| &= \left| \left\langle \varphi_{\sigma^2} * \left((uu^* - vv^*) \circ \boldsymbol{\tau}_{\nu_\natural - \nu} \right), \left\langle \nabla_{\boldsymbol{x}} [\varphi_{\sigma^2} * X_\natural], \boldsymbol{\tau}_{\pi/2} \right\rangle_{\ell^2} \right\rangle_{L^2(\mathbb{R}^2)} \right| \\
&\leq \left\| \varphi_{\sigma^2} * \left((uu^* - vv^*) \circ \boldsymbol{\tau}_{\nu_\natural - \nu} \right) \right\|_{L^2(\mathbb{R}^2)} \left\| \left\langle \nabla_{\boldsymbol{x}} [\varphi_{\sigma^2} * X_\natural], \boldsymbol{\tau}_{\pi/2} \right\rangle_{\ell^2} \right\|_{L^2(\mathbb{R}^2)}.
\end{aligned}$$

The second factor can be controlled with the second conclusion of Lemma F.13: this gives

$$\left\| \left\langle \nabla_{\boldsymbol{x}} [\varphi_{\sigma^2} * X_\natural], \boldsymbol{\tau}_{\pi/2} \right\rangle_{\ell^2} \right\|_{L^2(\mathbb{R}^2)} \leq \frac{1}{2\sigma^2} \left(1 + \sigma^2\right)^{1/2} \leq \frac{1}{\sigma^2},$$

where the last bound worst-cases with our assumption on σ. For the first factor, we use Young's convolution inequality and the fact that $f \mapsto f \circ \boldsymbol{\tau}_{\nu_\natural - \nu}$ is a unitary transformation of $L^2(\mathbb{R}^2)$ to obtain

$$\begin{aligned}
\left\| \varphi_{\sigma^2} * ((uu^* - vv^*) \circ \boldsymbol{\tau}_{-\nu}) \right\|_{L^2(\mathbb{R}^2)} &\leq \left\| \varphi_{\sigma^2} \right\|_{L^1(\mathbb{R}^2)} \left\| (uu^* - vv^*) \circ \boldsymbol{\tau}_{\nu_\natural - \nu} \right\|_{L^2(\mathbb{R}^2)} \\
&\leq \left\| uu^* - vv^* \right\|_{L^2(\mathbb{R}^2)}.
\end{aligned}$$

We have

$$\begin{aligned}
\left\| uu^* - vv^* \right\|_{L^2(\mathbb{R}^2)} &= \left| \|u\|_{L^2(\mathbb{R})}^2 - \|v\|_{L^2(\mathbb{R})}^2 \right| \\
&\leq 2 \max\left\{ \|u\|_{L^2(\mathbb{R})}, \|v\|_{L^2(\mathbb{R})} \right\} \left| \|u\|_{L^2(\mathbb{R})} - \|v\|_{L^2(\mathbb{R})} \right| \\
&\leq 2 \max\left\{ \|u\|_{L^2(\mathbb{R})}, \|v\|_{L^2(\mathbb{R})} \right\} \|u - v\|_{L^2(\mathbb{R})},
\end{aligned}$$

where the two inequalities are both applications of the triangle inequality. Combining these estimates, we have shown

$$|\nabla_\nu \mathcal{L}^\sigma(\nu, u) - \nabla_\nu \mathcal{L}^\sigma(\nu, v)| \leq \frac{2}{\sigma^2} \max\left\{ \|u_{\text{rough}}\|_{L^2(\mathbb{R})}, \|\bar{u}_{\text{rough}}\|_{L^2(\mathbb{R})} \right\} \|\bar{u}_{\text{rough}} - u_{\text{rough}}\|_{L^2(\mathbb{R})},$$

and applying this in the context of our gradient bounds, we obtain

$$\begin{aligned}
\left| \left(\nabla_\nu \mathcal{L}^\sigma(\,\cdot\,, u_{\text{rough}})\right)^2 - \left(\nabla_\nu \mathcal{L}^\sigma(\,\cdot\,, \bar{u}_{\text{rough}})\right)^2 \right| &\leq \frac{8}{\sigma^4} \max\left\{ \|u_{\text{rough}}\|_{L^2(\mathbb{R})}^3, \|\bar{u}_{\text{rough}}\|_{L^2(\mathbb{R})}^3 \right\} \|\bar{u}_{\text{rough}} - u_{\text{rough}}\|_{L^2(\mathbb{R})} \\
&\leq \frac{64}{\sigma^4} \|\bar{u}_{\text{rough}}\|_{L^2(\mathbb{R})}^3 \|\bar{u}_{\text{rough}} - u_{\text{rough}}\|_{L^2(\mathbb{R})},
\end{aligned}$$

where the final inequality simplifies using the triangle inequality and the assumption $\|u_{\text{rough}} - \bar{u}_{\text{rough}}\|_{L^2(\mathbb{R})} \leq \|\bar{u}_{\text{rough}}\|_{L^2(\mathbb{R})}$, as we used before. This is precisely the second assertion.

□

F.2.2 Technical Lemmas

Lemma F.11. *Let $u = \mathbb{1}_{[-\alpha,\alpha]}$ for some $\alpha > 0$, and for some smoothing level $\sigma > 0$ consider the associated curl field*

$$\mathcal{C}(\boldsymbol{x}) = \left\langle \nabla_{\boldsymbol{x}} [\varphi_{\sigma^2} * uu^*](\boldsymbol{x}), \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} \boldsymbol{x} \right\rangle_{\ell^2}.$$

If $\sigma^2 \leq \frac{\alpha^2}{24}$, then for any $\boldsymbol{x} = (s, t)$ with $0 \leq t \leq s$, one has

$$\mathcal{C}(s, t) \geq 0.$$

Proof. The proof uses expressions obtained in the proof of Lemma F.13. Following (F.79), if we define

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\alpha}^{\alpha} e^{-\frac{(x-x')^2}{2\sigma^2}} \, \mathrm{d}x', \quad x \in \mathbb{R},$$

then we have

$$\mathcal{C}(s,t) = sf(s)f'(t) - tf(t)f'(s).$$

Moreover, note that $f > 0$, and by (F.80) one has $f'(x) = 0$ only if $x = 0$. To show the claim, it therefore suffices to show that

$$\frac{f'(t)}{tf(t)} \geq \frac{f'(s)}{sf(s)}, \quad s \geq t > 0.$$

Differentiating, this monotonicity condition becomes

$$s(f'(s))^2 + f(s)f'(s) - sf''(s)f(s) \geq 0, \quad s > 0. \tag{F.72}$$

After a change of coordinates, we have

$$f(s) = \frac{1}{\sqrt{\pi}} \int_{-\alpha/\sqrt{2}\sigma}^{\alpha/\sqrt{2}\sigma} e^{-(s-\sqrt{2}\sigma x)^2/2\sigma^2} \, \mathrm{d}x.$$

Write $A = \alpha/\sqrt{2}\sigma$, and define

$$g(s) = \int_{-A}^{A} e^{-(s-x)^2} \, \mathrm{d}x.$$

Noting that $\pi^{-1/2} g(s/\sqrt{2}\sigma) = f(s)$ and inspecting (F.72), we see that it suffices to show that g satisfies the differential inequality in (F.72). Using the fundamental theorem of calculus, we have

$$\begin{aligned} g'(s) &= e^{-(s+A)^2} - e^{-(s-A)^2} \\ &= -2e^{-s^2-A^2} \sinh(2sA), \end{aligned} \tag{F.73}$$

and by an additional straightforward differentiation

$$g''(s) = 4e^{-s^2-A^2}(s\sinh(2sA) - A\cosh(2sA)).$$

After substituting into (F.72) and cancelling some positive factors, the inequality to show becomes

$$2sA\sinh(2sA)\left(\frac{A^{-1}e^{-A^2}}{e^{s^2}g(s)}\right) + 2sA\coth(2sA) - (1+2s^2) \geq 0. \tag{F.74}$$

We will prove this bound in two regimes: first, for $0 < s \leq A$, then for $s > A$.

Small s. To start, we will develop a simple estimate for g. Notice that

$$\begin{aligned} e^{s^2} g(s) &= \int_{-A}^{A} e^{-x^2} e^{2sx} \, \mathrm{d}x \\ &\leq \int_{-A}^{A} e^{2sx} \, \mathrm{d}x \\ &= \frac{1}{s} \sinh(2As), \end{aligned}$$

so it suffices to show

$$2s^2 e^{-A^2} + 2sA\coth(2sA) - (1+2s^2) \geq 0.$$

For large A, the first term is sub-leading, and it suffices to simply show

$$2sA\coth(2sA) - (1+2s^2) \geq 0.$$

We will show this bound on the requisite interval in two steps, since $s \mapsto s\coth s$ does not have a globally-convergent power series representation at zero. First, we have from the power series representation the bound $2sA\coth 2sA \geq 1+\frac{(2sA)^2}{3} - \frac{(2sA)^4}{45}$ for all s; this bound is initially valid for $|2sA| < \pi$, then extended to all s by noticing that it is decreasing for $2sA \geq \pi$, whereas $2sA\coth 2sA$ is increasing for $s \geq 0$. With this bound, it suffices to show

$$\frac{2s^2}{3}\left(2A^2 - 3 - \frac{8A^4s^2}{15}\right) \geq 0,$$

which, when $A^2 \geq 6$, holds for all $0 \leq s \leq A^{-1}\sqrt{45/16}$. Next, notice that (F.74) can be written equivalently as

$$s\left(e^{-(s-A)^2} - e^{-(s+A)^2}\right) + g(s)\left(2sA\coth(2sA) - (1+2s^2)\right) \geq 0, \tag{F.75}$$

where the first term is nonnegative. Since $\tanh(x) \leq 1$ if $x \geq 0$, it then suffices to show

$$g(s)\left(2sA - (1+2s^2)\right) \geq 0.$$

The concave quadratic function $2sA - (1+2s^2)$ has its two roots at $\frac{A}{2} \pm \frac{\sqrt{A^2-1}}{2}$; using the inequality $\sqrt{1-x} \geq 1-x$ for $0 \leq x \leq 1$, it follows that these two roots are outside of the interval $[\frac{1}{2A}, A - \frac{1}{2A}]$, and hence $2sA-(1+2s^2)$ is nonnegative on this interval. Since $\sqrt{45/16} \geq \frac{1}{2}$, this establishes the inequality on $0 < s \leq A - \frac{1}{2A}$.

Finally, to demonstrate the inequality on $A - \frac{1}{2A} \leq s \leq A$, we return to the sufficient expression of (F.74) given above, as

$$s\left(e^{-(s-A)^2} - e^{-(s+A)^2}\right) + g(s)\left(2sA - (1+2s^2)\right) \geq 0,$$

and note again that the concave quadratic function $2sA-(1+2s^2)$ is maximized at $s = A/2$, hence is a decreasing function of s on this interval; so it suffices to show

$$s\left(e^{-(s-A)^2} - e^{-(s+A)^2}\right) - g(s) \geq 0,$$

From (F.73), it is clear that g is a decreasing function of s, so we can show

$$s\left(e^{-(s-A)^2} - e^{-(s+A)^2}\right) - g(A - \tfrac{1}{2A}) \geq 0.$$

Now, exploiting the fact that the parameter s in the definition of $g(s)$ is similar to a "mean" parameter for the gaussian integrand, we calculate

$$\begin{aligned} g(A-\tfrac{1}{2A}) &= \int_{-A}^{A} e^{-\left(x-(A-\frac{1}{2A})\right)^2}\,\mathrm{d}x \\ &= \int_{-A}^{A-\frac{1}{2A}} e^{-\left(x-(A-\frac{1}{2A})\right)^2}\,\mathrm{d}x + \int_{A-\frac{1}{2A}}^{A} e^{-\left(x-(A-\frac{1}{2A})\right)^2}\,\mathrm{d}x \\ &\leq \frac{\sqrt{\pi}}{2} + \frac{1}{2A}. \end{aligned}$$

The last line above worst-cases the value of the first integral (as it is no larger than half of the integral over $\mathbb{R}$, by symmetry), and uses a L^1-L^∞ bound to control the second. Meanwhile, using elementary inequalities and $A^2 \geq 12$ assumed previously, we have

$$\begin{aligned} s\left(e^{-(s-A)^2} - e^{-(s+A)^2}\right) &\geq \left(A - \frac{1}{2A}\right)\left(1 - (s-A)^2 - e^{-\left(2A-\frac{1}{2A}\right)^2}\right) \\ &\geq \left(A - \frac{1}{2A}\right)\left(1 - \frac{1}{4A^2} - e^{-3A^2}\right), \end{aligned}$$

so it suffices to show

$$A\left(1 - \frac{1}{2A^2}\right)\left(1 - \frac{1}{4A^2} - e^{-3A^2}\right) - \left(\frac{\sqrt{\pi}}{2} + \frac{1}{2A}\right) \geq 0,$$

which is evidently true for all $A^2 \geq 12$. This establishes the inequality on $0 < s \leq A$.

Large s. For this regime, we will again proceed in steps; first for $A \le s \le A + \frac{1}{3}$, then for $A + \frac{1}{3} \le s \le 2A$, then for $s \ge 2A$.

First, we will develop the bound for $A \le s \le A + c$, where $c > 0$ is a small absolute constant. Proceeding as above, but using now that when $s \ge A$ we have $2sA - (1 + 2s^2) = -(1 + 2s(s - A) \le 0$ so that we can leverage the bound $g(s) = \int_{s-A}^{s+A} e^{-x^2}\,\mathrm{d}x \le \int_0^{s+A} e^{-x^2}\,\mathrm{d}x \le \sqrt{\pi}/2$, we have that it suffices to show

$$A\left(e^{-c^2} - e^{-4A^2}\right) - \frac{\sqrt{\pi}}{2}\left(1 + 2c(A + c)\right) \ge 0,$$

which, after rearranging, is simply

$$A\left(e^{-c^2} - \sqrt{\pi}c\right) - Ae^{-4A^2} - \sqrt{\pi}\left(\tfrac{1}{2} + c^2\right) \ge 0.$$

We verify numerically that this holds for $c = \frac{1}{3}$ when $A^2 \ge 12$, as we have assumed.

Now we proceed for $s \ge A + \frac{1}{3}$. We will develop a sequence of refined upper bounds on $g(s)$. We have from a change of coordinates

$$g(s) = e^{-(s-A)^2}\int_0^{2A} e^{-x^2 - 2x(s-A)}\,\mathrm{d}x. \tag{F.76}$$

Our previous estimate amounts to controlling the integral via $e^{-x^2} \le 1$. We will improve over this estimate slightly by instead developing a piecewise linear upper bound for the concave function $x \mapsto -x^2 - 2x(s - A)$. Below, we will write $t = s - A$ for concision; by assumption $t \ge \frac{1}{3}$. For any $\varepsilon \ge 0$, by concavity, we have for all $x \in \mathbb{R}$

$$-x^2 - 2x(s - A) \le -2(t + \varepsilon)x + \varepsilon^2.$$

Elementary algebra shows that the "null" upper bound for $\varepsilon = 0$, that is $x \mapsto -2tx$, intersects with $x \mapsto -2(t + \varepsilon)x + \varepsilon^2$ at $x = \varepsilon/2$. Hence, if $0 \le \varepsilon \le 4A$, we can estimate the integral as

$$\begin{aligned}\int_0^{2A} e^{-x^2-2x(s-A)}\,\mathrm{d}x &\le \int_0^{\varepsilon/2} e^{-2tx}\,\mathrm{d}x + e^{\varepsilon^2}\int_{\varepsilon/2}^{2A} e^{-2(t+\varepsilon)x}\,\mathrm{d}x \\ &= \frac{1}{2t}\left(1 - e^{-t\varepsilon}\right) + \frac{1}{2(t+\varepsilon)}\left(e^{-\varepsilon t} - e^{\varepsilon^2}e^{-4(t+\varepsilon)A}\right) \\ &\le \frac{1}{2t}\left(1 - \frac{\varepsilon e^{-\varepsilon t}}{\varepsilon + t}\right).\end{aligned}$$

Choosing $\varepsilon = 1/t$, this bound implies

$$g(s) \le \frac{1}{2t}e^{-(s-A)^2}\left(1 - \frac{e^{-1}}{1 + t^2}\right),$$

and so it suffices to show

$$e^{-(s-A)^2}\left(s - \frac{1}{2(s-A)}\left(1 - \frac{e^{-1}}{1 + (s-A)^2}\right)\left(1 + 2s(s - A)\right)\right) - se^{-(s+A)^2} \ge 0.$$

After some cancellation, this reads equivalently

$$\frac{1}{2(s-A)}\left(\frac{1 + 2s(s-A)}{e(1 + (s-A)^2)} - 1\right) - se^{-4sA} \ge 0.$$

We will estimate the term in parenthesis. We have

$$\frac{1 + 2s(s-A)}{e(1 + (s-A)^2)} - 1 = \left(\frac{-(1 - \frac{1}{e}) - (s - A)\left(s(1 - \frac{2}{e}) - A\right)}{1 + (s-A)^2}\right).$$

The numerator is a concave quadratic, which is maximized at $s = A\frac{e-1}{e-2}$; the constant is between 2 and 3. We check that when $s = A + \frac{1}{3}$ and $A^2 \geq 12$, the numerator is positive. Using $A + \frac{1}{3} \leq s \leq 2A$, we thus have that it suffices to show

$$\left(\frac{11 + 6A - 10e}{9e(1 + A^2)}\right) - \frac{4A}{3}e^{-4A^2} \geq 0.$$

We numerically verify that this inequality holds for all $A^2 \geq 12$.

Finally, we improve (F.76) once more and then use $s \geq 2A$ to conclude quickly. An improved estimate comes from [23, Theorem 1]: we apply this to (F.76) to obtain

$$\begin{aligned} g(s) &\leq e^{-(s-A)^2} \int_0^\infty e^{-x^2-2x(s-A)}\,\mathrm{d}x \\ &= \frac{\sqrt{\pi}}{2}\mathrm{erfc}(s-A) \\ &\leq \frac{e^{-(s-A)^2}}{2(s-A)}\left[1 - \frac{2 - 3e^{-(1+2(s-A))} - 2(s-A)e^{-(1+2(s-A))}}{(1+2(s-A))^2}\right], \end{aligned}$$

where the estimate is applied in the third line, and the second line is a standard integral. Following then (F.75) and using again that $|\tanh| \leq 1$, it suffices to show

$$1 - \frac{2 - 3e^{-(1+2(s-A))} - 2(s-A)e^{-(1+2(s-A))}}{(1+2(s-A))^2} \leq \frac{2s(s-A)\left(1 - e^{-4sA}\right)}{1 + 2s(s-A)}.$$

After rearranging with some algebra, it suffices to show

$$e^{-(1+2(s-A))}\frac{3 + 2(s-A)}{(1+2(s-A))^2} + e^{-4sA}\frac{2s(s-A)}{1+2s(s-A)} \leq \frac{2}{(1+2(s-A))^2} - \frac{1}{1+2s(s-A)}.$$

By algebra,

$$\frac{2}{(1+2(s-A))^2} - \frac{1}{1+2s(s-A)} = \frac{1 + 4sA - 4(s-A) - 4A^2}{(1+2s(s-A))(1+2(s-A))^2},$$

which is easily seen to be nonnegative when $s \geq A$. Clearing denominators, it then suffices to show

$$(3 + 2(s-A))(1 + 2s(s-A))e^{-(1+2(s-A))} + 2s(s-A)(1+2(s-A))^2 e^{-4sA} \leq 1 + 4sA - 4(s-A) - 4A^2.$$

We can show this holds easily by worst-casing for convenience, since the LHS has exponential prefactors. Since $s \geq 2A$, we have $2(s-A) \geq s$. We always have $s - A \leq s$, and since $A \geq 1$ we have $s \geq 2$, so it suffices to show

$$9s^3e^{-s} + 18s^4e^{-4sA} \leq 1 + 4sA - 4(s-A) - 4A^2.$$

Elementary calculus implies that the first term on the LHS is decreasing as soon as $s \geq 3$, and the second term is decreasing as soon as $s \geq 1/A$, both of which are implied by $s \geq 2A$ and our assumptions on A. Since the RHS is increasing, it suffices to check

$$9A^3e^{-A} + 18A^4e^{-4A^2} \leq 1 + 4A(A-1).$$

A numerical evaluation and the preceding calculus argument shows that this is true as soon as $A \geq 3$.

□

Lemma F.12. *Consider the residual field arising in the study of the alignment gradient: for any $\nu \in \mathbb{R}$, we consider*

$$\bar{u}_{\text{rough}}\bar{u}^*_{\text{rough}} \circ \boldsymbol{\tau}_\nu - \bar{u}_{\text{rough}}\bar{u}^*_{\text{rough}} \circ \boldsymbol{\tau}_{-\nu},$$

where the (unscaled, for convenience) nominal rough initial representation is defined as

$$\bar{u}_{\text{rough}}(s) = \mathbb{1}_{|s|\leq 1}\cos(\pi s/2).$$

Then for any $\boldsymbol{x} = (s,t)$ with $0 \le t \le s \le 1$, the difference is nonnegative:

$$\bar{u}_{\text{rough}}\bar{u}^*_{\text{rough}} \circ \boldsymbol{\tau}_\nu(\boldsymbol{x}) - \bar{u}_{\text{rough}}\bar{u}^*_{\text{rough}} \circ \boldsymbol{\tau}_{-\nu}(\boldsymbol{x}) \ge 0,$$

and moreover for any $\boldsymbol{x} = (s,t)$ with $0 \le t \le s \le 1$ and any $0 \le \nu \le \pi/7$, it satisfies the estimate

$$\bar{u}_{\text{rough}}\bar{u}^*_{\text{rough}} \circ \boldsymbol{\tau}_\nu(\boldsymbol{x}) - \bar{u}_{\text{rough}}\bar{u}^*_{\text{rough}} \circ \boldsymbol{\tau}_{-\nu}(\boldsymbol{x}) \ge \frac{7\sin\nu}{1000} \mathbb{1}_{-0.137 \le t - \frac{1}{\sqrt{2}} \le -0.127} \mathbb{1}_{-0.001 \le s - \frac{1}{\sqrt{2}} \le 0.001}.$$

Proof. We have to show

$$\bar{u}_{\text{rough}}\bar{u}^*_{\text{rough}} \circ \boldsymbol{\tau}_{-\nu} - \bar{u}_{\text{rough}}\bar{u}^*_{\text{rough}} \circ \boldsymbol{\tau}_{\nu} \le 0.$$

Using two trigonometric identities, we can write for any $\boldsymbol{x} = (s,t)$ with $0 \le t \le s \le 1$

$$\begin{aligned}
&\bar{u}_{\text{rough}}\bar{u}^*_{\text{rough}} \circ \boldsymbol{\tau}_{-\nu}(\boldsymbol{x}) - \bar{u}_{\text{rough}}\bar{u}^*_{\text{rough}} \circ \boldsymbol{\tau}_{\nu}(\boldsymbol{x}) \\
&\quad = \sin(\tfrac{\pi}{2}(s+t)\cos\nu)\sin(\tfrac{\pi}{2}(s-t)\sin\nu) - \sin(\tfrac{\pi}{2}(s-t)\cos\nu)\sin(\tfrac{\pi}{2}(s+t)\sin\nu).
\end{aligned}$$

It is clear that this expression is identically zero when $\nu = 0$ or $s = t$, so assume otherwise below. To show the expression is nonpositive, it is equivalent to show

$$\frac{\sin(\frac{\pi}{2}(s+t)\cos\nu)}{\sin(\frac{\pi}{2}(s+t)\sin\nu)} \le \frac{\sin(\frac{\pi}{2}(s-t)\cos\nu)}{\sin(\frac{\pi}{2}(s-t)\sin\nu)}$$

for each $0 \le t \le s \le 1$ and all $0 \le \nu \le \pi/4$. Given that under these constraints $s + t \le 2$, it therefore suffices to show that

$$x \mapsto \frac{\sin(\frac{\pi}{2}x\cos\nu)}{\sin(\frac{\pi}{2}x\sin\nu)}$$

is a decreasing function of x on $[0,2]$. Rescaling coordinates, this is equivalent to showing that

$$x \mapsto \frac{\sin(x\cot\nu)}{\sin x}$$

is decreasing on $[0, \pi\sin\nu]$, and because $1/\cot\nu \ge \sin\nu$ when $0 \le \nu \le \pi/4$ it suffices instead to show decreasingness on $[0, \pi/\cot\nu]$. This is a standard calculation that arises in the study of the Dirichlet kernel in Fourier analysis; to obtain it, write $A = \cot\nu$ and differentiate to obtain the sufficient condition

$$Ax\cot Ax \le x\cot x, \quad 0 < x < \pi/A.$$

This can be seen, for instance, from the power series expression for $x\cot x$, convergent for $|x| < \pi$:

$$Ax\cot Ax = 1 - 2\sum_{k=1}^{\infty} \frac{\zeta(2k)}{\pi^{2k}} A^{2k}x^{2k} \le 1 - 2\sum_{k=1}^{\infty} \frac{\zeta(2k)}{\pi^{2k}} x^{2k} = x\cot x,$$

since $A \ge 1$ and all terms in the sum are nonpositive, where ζ denotes the Riemann zeta function. Thus we have shown

$$\bar{u}_{\text{rough}}\bar{u}^*_{\text{rough}} \circ \boldsymbol{\tau}_{-\nu} - \bar{u}_{\text{rough}}\bar{u}^*_{\text{rough}} \circ \boldsymbol{\tau}_{\nu} \le 0.$$

Next, we show the quantitative bound. By our earlier work, we can write at any $\boldsymbol{x} = (s,t)$ with $0 \le t < s \le 1$

$$\begin{aligned}
&\frac{\sqrt{\pi}}{2}\left(\bar{u}_{\text{rough}}\bar{u}^*_{\text{rough}} \circ \boldsymbol{\tau}_{\nu}(\boldsymbol{x}) - \bar{u}_{\text{rough}}\bar{u}^*_{\text{rough}} \circ \boldsymbol{\tau}_{-\nu}(\boldsymbol{x})\right) \\
&\quad = \sin(\tfrac{\pi}{2}(s+t)\sin\nu)\sin(\tfrac{\pi}{2}(s-t)\cos\nu) - \sin(\tfrac{\pi}{2}(s+t)\cos\nu)\sin(\tfrac{\pi}{2}(s-t)\sin\nu). \qquad \text{(F.77)}
\end{aligned}$$

We will obtain a lower bound for this expression by combining term-by-term bounds, optimized for $t \approx s$. The difference terms are easiest: we can use the standard estimates

$$\begin{aligned}
\sin\left(\tfrac{\pi}{2}(s-t)\cos\nu\right) &\ge \tfrac{\pi}{2}(s-t)\cos\nu - \left(\tfrac{\pi}{2}(s-t)\cos\nu\right)^3/6, \\
\sin\left(\tfrac{\pi}{2}(s-t)\sin\nu\right) &\le \tfrac{\pi}{2}(s-t)\sin\nu,
\end{aligned}$$

which follow by concavity. Similarly, concavity gives the estimates

$$\sin\left(\tfrac{\pi}{2}(s+t)\sin\nu\right) \geq \sin\left(\pi s\sin\nu\right) + \tfrac{\pi}{2}\sin\nu\cos\left(\pi s\sin\nu\right)(t-s) - \frac{\pi^2\sin^2\nu}{8}(t-s)^2,$$
$$\sin\left(\tfrac{\pi}{2}(s+t)\cos\nu\right) \leq \sin\left(\pi s\cos\nu\right) + \tfrac{\pi}{2}\cos\nu\cos\left(\pi s\cos\nu\right)(t-s).$$

These estimates yield a polynomial lower bound for the difference term when substituted into (F.77). Since we know this difference is nonnegative, we can improve the bound by taking the maximum of it and zero, and then further simplify the bound based on its local behavior near $t \approx s$ to a quadratic lower bound. To this end, we have the unwieldy lower bound

$$\begin{aligned}\bar{u}_{\text{rough}}\bar{u}^*_{\text{rough}} \circ \boldsymbol{\tau}_\nu(\boldsymbol{x}) - \bar{u}_{\text{rough}}\bar{u}^*_{\text{rough}} \circ \boldsymbol{\tau}_{-\nu}(\boldsymbol{x}) \geq{}& \frac{\pi(s-t)}{2}\left(\cos(\nu)\sin(\pi s\sin\nu) - \sin(\nu)\sin(\pi s\cos\nu)\right)\\ &- \frac{\pi^2(s-t)^2\cos\nu\sin\nu}{4}\left(\cos(\pi s\sin\nu) - \cos(\pi s\cos\nu)\right)\\ &- \frac{\pi^3(s-t)^3\cos\nu}{16}\left(\sin^2\nu + \frac{1}{3}\cos^2\nu\sin(\pi s\sin\nu)\right)\\ &+ \frac{\pi^4(s-t)^4\cos^3(\nu)\cos(\pi s\sin\nu)\sin\nu}{96} + \frac{\pi^5(s-t)^5\cos^3(\nu)\sin^2(\nu)}{384}.\end{aligned}$$

The degree four and five terms in this bound are both nonnegative, hence can be worst-cased out. To verify that the degree-two term dominates the degree-three term, we have to show for some $0 < \varepsilon < 1$

$$(1-\varepsilon)\sin\nu\left(\cos(\pi s\sin\nu) - \cos(\pi s\cos\nu)\right) - \frac{\pi(s-t)}{4}\left(\sin^2\nu + \frac{1}{3}\cos^2\nu\sin(\pi s\sin\nu)\right) \geq 0.$$

We have $0 \leq s-t \leq 1$, and the LHS of the previous bound is a decreasing function of $s-t$. Moreover, we have $\sin(\pi s\sin\nu)/\sin\nu \leq \pi s$. Hence, to show this bound holds for all $0 \leq \nu \leq \pi/7$ and all $0 \leq s-t \leq 1/2$, it suffices to show for some ε

$$(1-\varepsilon)\left(\cos(\pi s\sin\nu) - \cos(\pi s\cos\nu)\right) - \frac{\pi}{8}\left(\sin\pi/7 + \frac{\pi s}{3}\right) \geq 0.$$

We will show this by lower bounding the first term with calculus. We have for the second derivative of the first summand

$$\partial_\nu^2[\cos(\pi s\sin\nu)](\nu) = \sin(\pi s\sin\nu)(\pi s\sin\nu) - (\pi s\cos\nu)^2\cos(\pi s\sin\nu).$$

We notice that this is an increasing function of ν, because it is a difference of two terms which are (respectively) a product of two nonnegative increasing functions and a product of two nonnegative decreasing functions. Hence it attains its minimum value at $\nu = 0$. Similarly, we have

$$\partial_\nu^2[-\cos(\pi s\cos\nu)](\nu) = (\pi s\sin\nu)^2\cos(\pi s - (\pi s\cos\nu)\sin(\pi s\cos\nu)\cos\nu),$$

which is once again a difference of a product of two nonnegative increasing functions and a product of two nonnegative decreasing functions, hence attains its minimum value at $\nu = 0$. It follows that the same is true of the sum, and Taylor's theorem then implies the lower bound

$$\cos(\pi s\sin\nu) - \cos(\pi s\cos\nu) \geq (1-\cos\pi s) - \frac{(\pi s)^2 + \pi s\sin\pi s}{2}\nu^2.$$

When $s \geq \frac{1}{2}$, we have $-\cos\pi s \geq 2s-1$. Worst-casing $\nu \leq \pi/7$ and choosing $\varepsilon = 1/8$, it then suffices to show under these conditions

$$\left(2s - \frac{(\pi s)^2 + \pi s\sin\pi s}{2}(\pi/7)^2\right) - \frac{\pi}{7}\left(\sin\pi/7 + \frac{\pi s}{3}\right) \geq 0.$$

A numerical evaluation shows that this holds for all $\frac{1}{2} \leq s \leq 1$. Hence, we have the lower bound, valid for $\frac{1}{2} \leq s \leq 1$, all $s - \frac{1}{2} \leq t \leq s$, and all $0 \leq \nu \leq \pi/7$:

$$\begin{aligned}\bar{u}_{\text{rough}}\bar{u}^*_{\text{rough}} \circ \boldsymbol{\tau}_\nu(\boldsymbol{x}) - \bar{u}_{\text{rough}}\bar{u}^*_{\text{rough}} \circ \boldsymbol{\tau}_{-\nu}(\boldsymbol{x}) \geq{}& \frac{\pi(s-t)}{2}\left(\cos(\nu)\sin(\pi s\sin\nu) - \sin(\nu)\sin(\pi s\cos\nu)\right)\\ &- \frac{15\pi^2(s-t)^2\cos\nu\sin\nu}{32}\left(\cos(\pi s\sin\nu) - \cos(\pi s\cos\nu)\right).\end{aligned}$$

Letting

$$A = \frac{\pi}{2}\left(\cos(\nu)\sin(\pi s\sin\nu) - \sin(\nu)\sin(\pi s\cos\nu)\right),$$
$$B = \frac{15\pi^2\cos\nu\sin\nu}{32}\left(\cos(\pi s\sin\nu) - \cos(\pi s\cos\nu)\right),$$

we have $B \geq 0$ since cos is decreasing and sin $\leq$ cos on our interval of interest, and

$$\bar{u}_{\text{rough}}\bar{u}^*_{\text{rough}} \circ \boldsymbol{\tau}_\nu(\boldsymbol{x}) - \bar{u}_{\text{rough}}\bar{u}^*_{\text{rough}} \circ \boldsymbol{\tau}_{-\nu}(\boldsymbol{x}) \geq Ar - Br^2 = B\left(\left(\frac{A}{2B}\right)^2 - \left(r - \frac{A}{2B}\right)^2\right),$$

where the RHS is a concave quadratic function of $r = s - t$. For such a concave quadratic, the above forms make it clear that its two roots are at 0 and A/B, and by concavity we have

$$Ar - Br^2 \geq \frac{3A^2}{16B}\mathbb{1}_{|r - A/2B| \leq |A/4B|}.$$

We will show that this bound also applies to $\bar{u}_{\text{rough}}\bar{u}^*_{\text{rough}} \circ \boldsymbol{\tau}_\nu(\boldsymbol{x}) - \bar{u}_{\text{rough}}\bar{u}^*_{\text{rough}} \circ \boldsymbol{\tau}_{-\nu}(\boldsymbol{x})$ on our interval of interest, by showing that $A \geq 0$ and when $A/4B \leq s - t \leq 3A/4B$, s and t satisfy the previously assumed conditions uniformly in ν. To see that $A \geq 0$, notice that

$$A = \frac{\pi\sin\nu\cos\nu}{2}\left(\frac{\sin(\pi s\sin\nu)}{\sin\nu} - \frac{\sin(\pi s\cos\nu)}{\cos\nu}\right).$$

The function $x \mapsto \sin x/x$ is decreasing when $0 \leq x \leq \pi$, showing that $A \geq 0$. This means that any s, t for which $s - t \geq A/4B$ satisfies our hypotheses. Next, note that

$$\nu \mapsto \frac{\sin(\pi s\sin\nu)}{\sin\nu} - \frac{\sin(\pi s\cos\nu)}{\cos\nu}$$

is decreasing, as $x \mapsto \sin x/x$ is nonnegative and decreasing for $0 \leq x \leq \pi$ (the chain rule implies the composition is decreasing as a sum of decreasing functions). By the same token,

$$\nu \mapsto \cos(\pi s\sin\nu) - \cos(\pi s\cos\nu)$$

is decreasing on our domain of interest. This implies

$$\frac{\frac{\sin(\pi s\sin\pi/7)}{\sin\pi/7} - \frac{\sin(\pi s\cos\pi/7)}{\cos\pi/7}}{1 - \cos(\pi s)} \leq \frac{\frac{\sin(\pi s\sin\nu)}{\sin\nu} - \frac{\sin(\pi s\cos\nu)}{\cos\nu}}{\cos(\pi s\sin\nu) - \cos(\pi s\cos\nu)} \leq \frac{\pi s - \sin(\pi s)}{\cos(\pi s\sin\pi/7) - \cos(\pi s\cos\pi/7)}.$$

A numerical evaluation shows that both the LHS and the RHS are increasing. Hence, if $s \leq 0.72$, we have the bound

$$\frac{\frac{\sin(\pi s\sin\nu)}{\sin\nu} - \frac{\sin(\pi s\cos\nu)}{\cos\nu}}{\cos(\pi s\sin\nu) - \cos(\pi s\cos\nu)} \leq \frac{0.72\pi - \sin(0.72\pi)}{\cos(0.72\pi\sin\pi/7) - \cos(0.72\pi\cos\pi/7)} \leq 1.483,$$

which implies

$$\frac{3A}{4B} \leq \frac{1.483\cdot 4}{5\pi} \leq 0.378,$$

showing that any s, t for which $s - t \leq 3A/4B$ satisfies our hypotheses. Finally, as above, using $s \geq 0.7$ we can obtain the lower bound

$$\frac{\frac{\sin(\pi s\sin\nu)}{\sin\nu} - \frac{\sin(\pi s\cos\nu)}{\cos\nu}}{\cos(\pi s\sin\nu) - \cos(\pi s\cos\nu)} \geq \frac{\frac{\sin(0.7\pi\sin\pi/7)}{\sin\pi/7} - \frac{\sin(0.7\pi\cos\pi/7)}{\cos\pi/7}}{1 - \cos(0.7\pi)} \geq 0.544,$$

which implies

$$\frac{A}{4B} \geq \frac{0.544\cdot 4}{15\pi} \geq 0.046.$$

Consequently, we have established

$$\bar{u}_{\text{rough}}\bar{u}^*_{\text{rough}} \circ \boldsymbol{\tau}_\nu(\boldsymbol{x}) - \bar{u}_{\text{rough}}\bar{u}^*_{\text{rough}} \circ \boldsymbol{\tau}_{-\nu}(\boldsymbol{x}) \geq \frac{9\pi(\cos(\nu)\sin(\pi s \sin\nu) - \sin(\nu)\sin(\pi s \cos\nu))}{800} \mathbb{1}_{|(s-t)-A/2B| \leq |A/4B|}.$$

As above, we can worst-case this bound further. Since

$$\begin{aligned}
\cos(\nu)\sin(\pi s \sin\nu) - \sin(\nu)\sin(\pi s\cos\nu) &= \sin(\nu)\cos(\nu)\left(\frac{\sin(\pi s \sin\nu)}{\sin\nu} - \frac{\sin(\pi s\cos\nu)}{\cos\nu}\right) \\
&\geq \sin(\nu)\cos(\pi/7)\left(\frac{\sin(\pi s \sin\pi/7)}{\sin\pi/7} - \frac{\sin(\pi s\cos\pi/7)}{\cos\pi/7}\right) \\
&\geq \sin(\nu)\cos(\pi/7)\left(\frac{\sin(\pi/2 \sin\pi/7)}{\sin\pi/7} - \frac{\sin(\pi/2\cos\pi/7)}{\cos\pi/7}\right) \\
&\geq \frac{\sin\nu}{5},
\end{aligned}$$

we have

$$\bar{u}_{\text{rough}}\bar{u}^*_{\text{rough}} \circ \boldsymbol{\tau}_\nu(\boldsymbol{x}) - \bar{u}_{\text{rough}}\bar{u}^*_{\text{rough}} \circ \boldsymbol{\tau}_{-\nu}(\boldsymbol{x}) \geq \frac{7\sin\nu}{1000} \mathbb{1}_{|(s-t)-A/2B| \leq |A/4B|}.$$

In addition, we have shown above

$$0.184 \leq \frac{A}{B} \leq 0.504,$$

which implies

$$\frac{A}{4B} \leq 0.126, \quad \frac{3A}{4B} \geq 0.138,$$

whence

$$\bar{u}_{\text{rough}}\bar{u}^*_{\text{rough}} \circ \boldsymbol{\tau}_\nu(\boldsymbol{x}) - \bar{u}_{\text{rough}}\bar{u}^*_{\text{rough}} \circ \boldsymbol{\tau}_{-\nu}(\boldsymbol{x}) \geq \frac{7\sin\nu}{1000} \mathbb{1}_{0.126 \leq (s-t) \leq 0.138}.$$

Because we have shown that the LHS is nonnegative previously, this bound holds for all t. Now notice that we can write the constraint on t on the RHS equivalently as

$$0.126 \leq s - t \leq 0.138 \iff (s - 0.132) - 0.006 \leq t \leq (s - 0.132) + 0.006.$$

Hence, if we consider a sub-interval of valid s, namely $s \in [\frac{1}{\sqrt{2}} - 0.001, \frac{1}{\sqrt{2}} + 0.001]$, we have for such s

$$0.126 \leq s - t \leq 0.138 \impliedby (\tfrac{1}{\sqrt{2}} - 0.132) - 0.005 \leq t \leq (\tfrac{1}{\sqrt{2}} - 0.132) + 0.005.$$

In particular,

$$\mathbb{1}_{0.126 \leq (s-t) \leq 0.138} \mathbb{1}_{0.7 \leq s \leq 0.72} \geq \mathbb{1}_{-0.137 \leq t - \frac{1}{\sqrt{2}} \leq -0.127} \mathbb{1}_{-0.001 \leq s - \frac{1}{\sqrt{2}} \leq 0.001}.$$

□

Lemma F.13. *For $\beta > 0$, let $u = \mathbb{1}_{[-\beta,\beta]}$, and for some smoothing level $\sigma > 0$ consider the associated curl fields*

$$\mathcal{C}^\beta(\boldsymbol{x}) = \left\langle \nabla_{\boldsymbol{x}}[\varphi_{\sigma^2} * uu^*](\boldsymbol{x}), \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} \boldsymbol{x} \right\rangle_{\ell^2}. \tag{F.78}$$

We have the following estimates: if $\sigma^2 \geq 1$ and $1/\sqrt{2} \leq \alpha \leq 1$, then

$$\langle \mathcal{C}^1, \mathcal{C}^\alpha \rangle_{L^2(\mathbb{R}^2)} \geq \frac{1}{8\pi\sigma^4},$$

and for any β and any $\sigma^2 > 0$,

$$\langle \mathcal{C}^\beta, \mathcal{C}^\beta \rangle_{L^2(\mathbb{R}^2)} \leq \frac{\beta^4}{\sigma^4}\left(\sigma^2 + 2\beta^2\right).$$

Proof. If we unravel the expression (F.78), we have

$$\begin{aligned}\mathcal{C}(s,t) &= \langle \nabla_{\boldsymbol{x}}[\varphi_{\sigma^2} * uu^*](s,t), (-t,s)\rangle_{\ell^2} \\ &= \left\langle \int_{\mathbb{R}}\int_{\mathbb{R}} \begin{bmatrix} u(s')u(t')\varphi_{\sigma^2}(t-t')\nabla_s[\varphi_{\sigma^2}](s-s') \\ u(s')u(t')\varphi_{\sigma^2}(s-s')\nabla_t[\varphi_{\sigma^2}](t-t') \end{bmatrix} \mathrm{d}s\,\mathrm{d}t, (-t,s) \right\rangle_{\ell^2} \\ &= \mathbf{1}^* \begin{bmatrix} -t\left(\int_{\mathbb{R}} u(s')\nabla_s\varphi_{\sigma^2}(s-s')\,\mathrm{d}s'\right)\left(\int_{\mathbb{R}} u(t')\varphi_{\sigma^2}(t-t')\,\mathrm{d}t'\right) \\ s\left(\int_{\mathbb{R}} u(s')\varphi_{\sigma^2}(s-s')\,\mathrm{d}s'\right)\left(\int_{\mathbb{R}} u(t')\nabla_t\varphi_{\sigma^2}(t-t')\,\mathrm{d}t'\right) \end{bmatrix}.\end{aligned}$$

Defining

$$\begin{aligned} f_1(x) &= \int_{\mathbb{R}} u(x')\nabla\varphi_{\sigma^2}(x-x')\,\mathrm{d}x'; \\ f_2(x) &= -x\int_{\mathbb{R}} u(x')\varphi_{\sigma^2}(x-x')\,\mathrm{d}x', \end{aligned}$$

the above implies

$$\mathcal{C}(s,t) = f_1(s)f_2(t) - f_1(t)f_2(s). \tag{F.79}$$

Now notice that, by the fundamental theorem of calculus,

$$\begin{aligned} f_1(x) &= \int_{\mathbb{R}} u(x-x')\nabla\varphi_{\sigma^2}(x')\,\mathrm{d}x' \\ &= \int_{x-\beta}^{x+\beta} \nabla\varphi_{\sigma^2}(x')\,\mathrm{d}x' \\ &= \varphi_{\sigma^2}(x+\beta) - \varphi_{\sigma^2}(x-\beta). \end{aligned} \tag{F.80}$$

For the estimates we need, we introduce more general notation: for any $\gamma > 0$ (following (F.80)), let

$$\begin{aligned} f_1^\gamma(x) &= \varphi_{\sigma^2}(x+\gamma) - \varphi_{\sigma^2}(x-\gamma); \\ f_2^\gamma(x) &= -x\int_{\mathbb{R}} \mathbb{1}_{[-\gamma,\gamma]}(x')\varphi_{\sigma^2}(x-x')\,\mathrm{d}x'. \end{aligned}$$

Our task is to estimate Ξ, defined as

$$\Xi(\alpha,\beta) = \left\langle f_1^\alpha (f_2^\alpha)^* - f_2^\alpha (f_1^\alpha)^*, f_1^\beta (f_2^\beta)^* - f_2^\beta (f_1^\beta)^* \right\rangle_{L^2(\mathbb{R}^2)},$$

since by the analysis above we have $\langle \mathcal{C}^\alpha, \mathcal{C}^\beta\rangle_{L^2(\mathbb{R}^2)} = \Xi(\alpha,\beta)$. Distributing in the inner product, we have

$$\Xi(\alpha,\beta) = 2\left(\langle f_1^\alpha, f_1^\beta\rangle\langle f_2^\alpha, f_2^\beta\rangle - \langle f_1^\alpha, f_2^\beta\rangle\langle f_2^\alpha, f_1^\beta\rangle\right). \tag{F.81}$$

We first estimate the cross terms $\langle f_2^\alpha, f_1^\beta\rangle$. We have by (F.80)

$$\langle f_2^\alpha, f_1^\beta\rangle = \int_{\mathbb{R}} x(\mathbb{1}_{[-\alpha,\alpha]} * \varphi_{\sigma^2})(x)\left(\varphi_{\sigma^2}(x-\beta) - \varphi_{\sigma^2}(x+\beta)\right)\mathrm{d}x.$$

As above, we can integrate this out once again. We have, following the argument in (F.80)

$$\begin{aligned} \int_{\mathbb{R}} x\mathbb{1}_{[-\alpha,\alpha]}(x-x')\varphi_{\sigma^2}(x-\beta)\,\mathrm{d}x &= \int_{\mathbb{R}} (x-\beta)\mathbb{1}_{[-\alpha,\alpha]}(x-x')\varphi_{\sigma^2}(x-\beta)\,\mathrm{d}x + \beta\int_{\mathbb{R}} \mathbb{1}_{[-\alpha,\alpha]}(x-x')\varphi_{\sigma^2}(x-\beta)\,\mathrm{d}x \\ &= -\sigma^2\int_{\mathbb{R}} \mathbb{1}_{[-\alpha,\alpha]}(x-x')\nabla\varphi_{\sigma^2}(x-\beta)\,\mathrm{d}x + \beta\int_{\mathbb{R}} \mathbb{1}_{[-\alpha,\alpha]}(x-x')\varphi_{\sigma^2}(x-\beta)\,\mathrm{d}x \\ &= \sigma^2\left(\varphi_{\sigma^2}(x'-\alpha-\beta) - \varphi_{\sigma^2}(x'+\alpha-\beta)\right) + \beta\int_{\mathbb{R}} \mathbb{1}_{[-\alpha,\alpha]}(x-x')\varphi_{\sigma^2}(x-\beta)\,\mathrm{d}x. \end{aligned}$$

Reasoning symmetrically, we get

$$\begin{aligned}\int_{\mathbb{R}} x\mathbb{1}_{[-\alpha,\alpha]}(x-x')\left(\varphi_{\sigma^2}(x-\beta)-\varphi_{\sigma^2}(x+\beta)\right)\mathrm{d}x &= \sigma^2\big(\varphi_{\sigma^2}(x'-\alpha-\beta)-\varphi_{\sigma^2}(x'+\alpha-\beta)\\ &\quad -\varphi_{\sigma^2}(x'-\alpha+\beta)+\varphi_{\sigma^2}(x'+\alpha+\beta)\big)\\ &\quad +\beta\int_{\mathbb{R}}\mathbb{1}_{[-\alpha,\alpha]}(x-x')\left(\varphi_{\sigma^2}(x-\beta)+\varphi_{\sigma^2}(x+\beta)\right)\mathrm{d}x.\end{aligned}$$

To obtain $\langle f_2^\alpha, f_1^\beta\rangle$ from this last expression, we integrate against $\varphi_{\sigma^2}(x')$. Integrating this function against the first term on the RHS of the previous expression yields a convolution between gaussians, which is another gaussian:

$$\begin{aligned}\sigma^2\int_{\mathbb{R}}\varphi_{\sigma^2}(x')\big(\varphi_{\sigma^2}(x'-\alpha-\beta)-\varphi_{\sigma^2}(x'+\alpha-\beta)-\varphi_{\sigma^2}(x'-\alpha+\beta)+\varphi_{\sigma^2}(x'+\alpha+\beta)\big)\,\mathrm{d}x'\\ = 2\sigma^2\left(\varphi_{2\sigma^2}(\alpha+\beta)-\varphi_{2\sigma^2}(\alpha-\beta)\right),\end{aligned}\tag{F.82}$$

where we used even symmetry of the gaussian. Integrating against the second term can be similarly manipulated to give

$$\begin{aligned}&\beta\int_{\mathbb{R}}\int_{\mathbb{R}}\mathbb{1}_{[-\alpha,\alpha]}(x-x')\left(\varphi_{\sigma^2}(x-\beta)+\varphi_{\sigma^2}(x+\beta)\right)\varphi_{\sigma^2}(x')\,\mathrm{d}x\\ &\quad=\beta\int_{\mathbb{R}}\int_{\mathbb{R}}\mathbb{1}_{[-\alpha,\alpha]}(x)\left(\varphi_{\sigma^2}(x+x'-\beta)+\varphi_{\sigma^2}(x+x'+\beta)\right)\varphi_{\sigma^2}(x')\,\mathrm{d}x\\ &\quad=\beta\int_{\mathbb{R}}\int_{\mathbb{R}}\mathbb{1}_{[-\alpha,\alpha]}(x)\left(\varphi_{\sigma^2}(\beta-x-x')+\varphi_{\sigma^2}(-\beta-x-x')\right)\varphi_{\sigma^2}(x')\,\mathrm{d}x\\ &\quad=\beta\int_{\mathbb{R}}\mathbb{1}_{[-\alpha,\alpha]}(x)\left(\varphi_{2\sigma^2}(\beta-x)+\varphi_{2\sigma^2}(-\beta-x)\right)\mathrm{d}x\\ &=\beta\left(\mathbb{1}_{[-\alpha,\alpha]}*\varphi_{2\sigma^2}(\beta)+\mathbb{1}_{[-\alpha,\alpha]}*\varphi_{2\sigma^2}(-\beta)\right).\end{aligned}$$

Thus

$$\langle f_2^\alpha, f_1^\beta\rangle = 2\sigma^2\left(\varphi_{2\sigma^2}(\alpha+\beta)-\varphi_{2\sigma^2}(\alpha-\beta)\right)+2\beta\mathbb{1}_{[-\alpha,\alpha]}*\varphi_{2\sigma^2}(\beta).$$

The remaining calculations proceed along similar lines. We have by symmetry

$$\begin{aligned}\langle f_1^\alpha, f_1^\beta\rangle &= 2\langle\varphi_{\sigma^2}(\,\cdot+\alpha),\varphi_{\sigma^2}(\,\cdot+\beta)\rangle_{L^2}-2\langle\varphi_{\sigma^2}(\,\cdot+\alpha),\varphi_{\sigma^2}(\,\cdot-\beta)\rangle_{L^2}\\ &= 2\left(\varphi_{2\sigma^2}(\beta-\alpha)-\varphi_{2\sigma^2}(\beta+\alpha)\right),\end{aligned}\tag{F.83}$$

because the integrals are gaussian convolutions. Notice that this implies

$$\langle f_2^\alpha, f_1^\beta\rangle = 2\beta(\mathbb{1}_{[-\alpha,\alpha]}*\varphi_{2\sigma^2})(\beta)-\sigma^2\langle f_1^\alpha, f_1^\beta\rangle.\tag{F.84}$$

For the remaining integral, we start with

$$\langle f_2^\alpha, f_2^\beta\rangle = \left\langle(\,\cdot\,)\int_{\mathbb{R}}\mathbb{1}_{[-\alpha,\alpha]}(x')\varphi_{\sigma^2}(\,\cdot-x')\,\mathrm{d}x',(\,\cdot\,)\int_{\mathbb{R}}\mathbb{1}_{[-\beta,\beta]}(x')\varphi_{\sigma^2}(\,\cdot-x')\,\mathrm{d}x'\right\rangle,$$

which motivates us to consider

$$\begin{aligned}\int_{\mathbb{R}}x^2\varphi_{\sigma^2}(x-x')\varphi_{\sigma^2}(x-x'')\,\mathrm{d}x &= \varphi_{2\sigma^2}(x'-x'')\int_{\mathbb{R}}x^2\varphi_{\sigma^2/2}\left(x-\tfrac{x'+x''}{2}\right)\mathrm{d}x\\ &= \varphi_{2\sigma^2}(x'-x'')\int_{\mathbb{R}}\left(x+\tfrac{x'+x''}{2}\right)^2\varphi_{\sigma^2/2}(x)\,\mathrm{d}x\\ &= \varphi_{2\sigma^2}(x'-x'')\left(\frac{\sigma^2}{2}+\frac{(x'+x'')^2}{4}\right),\end{aligned}$$

where the first line follows by completing the square. In particular, this shows that

$$\langle f_2^\alpha, f_2^\beta\rangle = \int_{\mathbb{R}}\int_{\mathbb{R}}\mathbb{1}_{[-\alpha,\alpha]}(x')\mathbb{1}_{[-\beta,\beta]}(x'')\varphi_{2\sigma^2}(x'-x'')\left(\frac{\sigma^2}{2}+\frac{(x'+x'')^2}{4}\right)\mathrm{d}x'\,\mathrm{d}x''.\tag{F.85}$$

We turn to using these calculations to obtain the remaining estimates. From (F.85), we have (because all terms in the integral are nonnegative)

$$\begin{aligned}
\langle f_2^\alpha, f_2^\beta \rangle &\geq \frac{\sigma^2}{2} \int_{\mathbb{R}} \int_{\mathbb{R}} \mathbb{1}_{[-\alpha,\alpha]}(x') \mathbb{1}_{[-\beta,\beta]}(x'') \varphi_{2\sigma^2}(x' - x'') \, \mathrm{d}x' \, \mathrm{d}x'' \\
&= \frac{\varphi_{2\sigma^2}(0)\sigma^2}{2} \int_{\mathbb{R}} \int_{\mathbb{R}} \mathbb{1}_{[-\alpha,\alpha]}(x') \mathbb{1}_{[-\beta,\beta]}(x'') \\
&\quad - \frac{\sigma^2}{2} \int_{\mathbb{R}} \int_{\mathbb{R}} \mathbb{1}_{[-\alpha,\alpha]}(x') \mathbb{1}_{[-\beta,\beta]}(x'') \left(\varphi_{2\sigma^2}(0) - \varphi_{2\sigma^2}(x' - x'') \right) \mathrm{d}x' \, \mathrm{d}x'' \\
&\geq 2\alpha\beta\sigma^2 \varphi_{2\sigma^2}(0) - \frac{\varphi_{2\sigma^2}(0)}{8} \int_{\mathbb{R}} \int_{\mathbb{R}} \mathbb{1}_{[-\alpha,\alpha]}(x') \mathbb{1}_{[-\beta,\beta]}(x'') (x' - x'')^2 \, \mathrm{d}x' \, \mathrm{d}x'' \\
&= 2\alpha\beta\sigma^2 \varphi_{2\sigma^2}(0) - \frac{\varphi_{2\sigma^2}(0)}{8} \int_{-\alpha}^{\alpha} \int_{-\beta}^{\beta} (x' - x'')^2 \, \mathrm{d}x' \, \mathrm{d}x'' \\
&= 2\alpha\beta\sigma^2 \varphi_{2\sigma^2}(0) - \frac{\varphi_{2\sigma^2}(0)}{6} \left(\alpha^3\beta + \beta^3\alpha \right), \qquad \text{(F.86)}
\end{aligned}$$

where the second line applies the triangle inequality, and the third uses the inequality $1 - e^{-x} \leq x$. From (F.84) and (F.83), we require upper and lower bounds on (F.83). We have

$$\langle f_1^\alpha, f_1^\beta \rangle = 2\varphi_{2\sigma^2}(0) e^{-\frac{1}{4\sigma^2}(\beta-\alpha)^2} \left(1 - e^{-\frac{1}{4\sigma^2}((\beta+\alpha)^2 - (\alpha-\beta)^2)} \right).$$

Upper bounds from here are straightforward, using that $e^{-x} \leq 1$ for $x \geq 0$ and $1 - e^{-x} \leq x$. We get

$$\langle f_1^\alpha, f_1^\beta \rangle \leq \frac{2\alpha\beta\varphi_{2\sigma^2}(0)}{\sigma^2}. \qquad \text{(F.87)}$$

Lower bounds can be obtained similarly: by the mean value theorem, there is a $\xi \in (1/(4\sigma^2))[(\beta-\alpha)^2, (\beta+\alpha)^2]$ such that

$$\begin{aligned}
\left(e^{-\frac{(\beta-\alpha)^2}{4\sigma^2}} - e^{-\frac{(\alpha+\beta)^2}{4\sigma^2}} \right) &= e^{-\xi} \left(\frac{(\alpha+\beta)^2}{4\sigma^2} - \frac{(\beta-\alpha)^2}{4\sigma^2} \right) \\
&= e^{-\xi} \frac{\alpha\beta}{\sigma^2}.
\end{aligned}$$

Using the lower bound on ξ and the fact that $e^{-x} \geq 1 - x$ gives the lower bound

$$\langle f_1^\alpha, f_1^\beta \rangle \geq \frac{2\alpha\beta\varphi_{2\sigma^2}(0)}{\sigma^2} \left(1 - \frac{(\beta-\alpha)^2}{4\sigma^2} \right). \qquad \text{(F.88)}$$

It remains to estimate the remaining term in (F.84). We write

$$\begin{aligned}
2\beta(\mathbb{1}_{[-\alpha,\alpha]} * \varphi_{2\sigma^2})(\beta) &= 4\alpha\beta \int_{\mathbb{R}} \frac{1}{2\alpha} \mathbb{1}_{[-\alpha,\alpha]}(x') \varphi_{2\sigma^2}(\beta - x') \, \mathrm{d}x' \\
&\leq 4\alpha\beta\varphi_{2\sigma^2}(0) \int_{\mathbb{R}} \frac{1}{2\alpha} \mathbb{1}_{[-\alpha,\alpha]}(x') \left(1 - \frac{(\beta-x')^2}{4\sigma^2} + \frac{(\beta-x')^4}{32\sigma^4} \right) \mathrm{d}x' \\
&= 4\alpha\beta\varphi_{2\sigma^2}(0) \left(1 - \frac{1}{4\sigma^2} \left(\frac{\alpha^2}{3} + \beta^2 \right) + \frac{1}{32\sigma^4} \left(\frac{\alpha^4}{5} + 2\alpha^2\beta^2 + \beta^4 \right) \right) \qquad \text{(F.89)}
\end{aligned}$$

using again $e^{-x} \leq 1 - x + \frac{1}{2}x^2$ in the second line. Plugging (F.87), (F.88), (F.89) and (F.86) into (F.81), we have the estimate

$$\begin{aligned}
\frac{1}{2}\Xi(\alpha,\beta) \geq & 4\alpha^2\beta^2 \left(\varphi_{2\sigma^2}(0) \right)^2 \left(\left(1 - \frac{(\beta-\alpha)^2}{4\sigma^2} \right) \left(1 - \frac{1}{12\sigma^2} \left(\alpha^2 + \beta^2 \right) \right) \right. \\
& \left. - \left(1 - \frac{1}{2\sigma^2} \left(\frac{\alpha^2}{3} + \beta^2 \right) + \frac{1}{16\sigma^4} \left(\frac{\alpha^4}{5} + 2\alpha^2\beta^2 + \beta^4 \right) \right) \left(1 - \frac{1}{2\sigma^2} \left(\frac{\beta^2}{3} + \alpha^2 \right) + \frac{1}{16\sigma^4} \left(\frac{\beta^4}{5} + 2\alpha^2\beta^2 + \alpha^4 \right) \right) \right),
\end{aligned}$$

where plugging in in this manner is justified by the fact that both factors in the product to the right of the minus sign are positive as long as $\sigma^2 \geq \alpha^2/6 + \beta^2/2$. Specializing to our setting of interest where $\alpha \leq 1$ and $\beta = 1$ and collecting terms makes this bound become (after simplifying constants numerically)

$$\frac{1}{2}\Xi(\alpha, 1) \geq \frac{1}{2\pi\sigma^4}\left(2/3 + \alpha/2 - \frac{0.845}{\sigma^2} - \frac{0.04}{\sigma^6}\right),$$

and the requirement is $\sigma^2 \geq 2/3$. Choosing $\sigma \geq 1$ and $\alpha \geq 1/\sqrt{2}$, the term in parentheses is no smaller than $1/8$, which gives the lower bound

$$\Xi(\alpha, 1) \geq \frac{1}{8\pi\sigma^4}.$$

The remaining upper bounds can be obtained easily from our work above. Notice that

$$\begin{aligned}\frac{1}{2}\Xi(\alpha, \alpha) &= \|f_1^\alpha\|_{L^2}^2\|f_2^\alpha\|_{L^2}^2 - \langle f_1^\alpha, f_2^\alpha\rangle^2 \\ &\leq \|f_1^\alpha\|_{L^2}^2\|f_2^\alpha\|_{L^2}^2.\end{aligned}$$

(F.87) gives a suitable upper bound on the first term; we only need to develop an upper bound on the second term. From (F.85), we proceed as

$$\begin{aligned}\langle f_2^\alpha, f_2^\alpha\rangle &= \int_{\mathbb{R}}\int_{\mathbb{R}} \mathbb{1}_{[-\alpha,\alpha]}(x')\mathbb{1}_{[-\alpha,\alpha]}(x'')\varphi_{2\sigma^2}(x' - x'')\left(\frac{\sigma^2}{2} + \frac{(x' + x'')^2}{4}\right)\mathrm{d}x'\,\mathrm{d}x'' \\ &\leq \int_{\{s^2+t^2\leq 2\alpha^2\}} \varphi_{2\sigma^2}(s - t)\left(\frac{\sigma^2}{2} + \frac{(s + t)^2}{4}\right)\mathrm{d}x'\,\mathrm{d}x'' \\ &\leq \int_{\{s^2+t^2\leq 2\alpha^2\}} \varphi_{2\sigma^2}(s - t)\left(\frac{\sigma^2}{2} + \alpha^2\right)\mathrm{d}x'\,\mathrm{d}x'' \\ &\leq \left(\frac{\sigma^2}{2} + \alpha^2\right)\int_{\{s^2+t^2\leq 2\alpha^2\}} \varphi_{2\sigma^2}(\sqrt{2}s)\,\mathrm{d}x'\,\mathrm{d}x'' \\ &\leq \varphi_{2\sigma^2}(0)\left(\frac{\sigma^2}{2} + \alpha^2\right)2\pi\alpha^2,\end{aligned}$$

where we pass to an enclosing circular domain in the second line by the fact that the integrand is nonnegative, use Cauchy-Schwarz in the third line and replace $(s+t)^2 \leq 2s^2+2t^2$ by its maximum over the domain of integration, apply an orthogonal change of coordinates in the fourth line, and use Hölder's inequality for the fifth line. Thus, invoking also (F.87), we have

$$\Xi(\alpha, \alpha) \leq \frac{\alpha^4}{\sigma^4}\left(\sigma^2 + 2\alpha^2\right).$$

□

Lemma F.14. *Let $u = \mathbb{1}_{[-\alpha,\alpha]}$ for some $\alpha > 0$, and for some smoothing level $\sigma > 0$ consider the associated curl field*

$$\mathcal{C}(\boldsymbol{x}) = \left\langle \nabla_{\boldsymbol{x}}[\varphi_{\sigma^2} * uu^*](\boldsymbol{x}), \begin{bmatrix}0 & -1\\ 1 & 0\end{bmatrix}\boldsymbol{x}\right\rangle_{\ell^2}.$$

Writing $\boldsymbol{x} = (s, t)$, and defining

$$f(s) = \int_{-\alpha}^{\alpha} \varphi_{\sigma^2}(s - x)\,\mathrm{d}x,$$

as in the proof of Lemma F.13, we have the explicit expression

$$\nabla_{\boldsymbol{x}}\mathcal{C}(\boldsymbol{x}) = \begin{bmatrix}sf'(s)f'(t) + f(s)f'(t) - tf(t)f''(s)\\ sf(s)f''(t) - f(t)f'(s) - tf'(s)f'(t)\end{bmatrix}, \tag{F.90}$$

and the 'iterated curl field' satisfies the estimate

$$\int_{\mathbb{R}^2} \left(\left\langle \nabla_{\boldsymbol{x}} \mathcal{C}(s,t), \begin{bmatrix} -t \\ s \end{bmatrix} \right\rangle_{\ell^2} \right)^2 \mathrm{d}s\,\mathrm{d}t \leq \frac{28\alpha^4}{\pi\sigma^2} + \frac{3\alpha^6(20\sigma^2 + 4\alpha^2)}{10\pi\sigma^6},$$

and if $\alpha^2 \leq 1$, it also satisfies the estimate (which is better when σ is small)

$$\int_{\mathbb{R}^2} \left(\left\langle \nabla_{\boldsymbol{x}} \mathcal{C}(s,t), \begin{bmatrix} -t \\ s \end{bmatrix} \right\rangle_{\ell^2} \right)^2 \mathrm{d}s\,\mathrm{d}t \leq \frac{3}{\pi} + \frac{55}{\pi\sigma^2} + \frac{4}{5\pi\sigma^4}.$$

Above, we use φ_{σ^2} interchangeably for a one-dimensional gaussian function and a two-dimensional gaussian function, with the meaning clear from the dimensionality of its argument.

Proof. Following the proof of Lemma F.13, we have

$$\mathcal{C}(s,t) = sf(s)f'(t) - tf(t)f'(s),$$

where by the fundamental theorem of calculus,

$$f'(x) = \varphi_{\sigma^2}(x+\alpha) - \varphi_{\sigma^2}(x-\alpha).$$

It is straightforward to calculate (F.90) from this expression. We have

$$\left\langle \nabla_{\boldsymbol{x}} \mathcal{C}(s,t), \begin{bmatrix} -t \\ s \end{bmatrix} \right\rangle_{\ell^2} = f(t)\left(t^2 f''(s) - sf'(s)\right) + f(s)\left(s^2 f''(t) - tf'(t)\right) - 2stf'(s)f'(t).$$

We square and integrate this expression in order to take care of the permutation symmetry. Using computer algebra software, one obtains

$$\begin{aligned}
\int_{\mathbb{R}^2} \left(\left\langle \nabla_{\boldsymbol{x}} \mathcal{C}(s,t), \begin{bmatrix} -t \\ s \end{bmatrix} \right\rangle_{\ell^2} \right)^2 \mathrm{d}s\,\mathrm{d}t &= 2\|p_x f'\|_{L^2}^2 \|f\|_{L^2}^2 + 8\|p_x f'\|_{L^2}^2 \int_{\mathbb{R}} p_x f f' + 2\|f''\|_{L^2}^2 \|p_{x^2} f\|_{L^2}^2 \\
&\quad - 4\left(\int_{\mathbb{R}} f f''\right)\left(\int_{\mathbb{R}} p_{x^3} f f'\right) - 8\left(\int_{\mathbb{R}} p_x f' f''\right)\left(\int_{\mathbb{R}} p_{x^3} f f'\right) \\
&\quad - 4\|p_x f\|_{L^2}^2 \int_{\mathbb{R}} p_x f' f'' + 2\left(\int_{\mathbb{R}} p_x f f'\right)^2 + 4\|p_x f'\|_{L^2}^4 + 2\left(\int_{\mathbb{R}} p_{x^2} f f''\right)^2.
\end{aligned}$$

In this expression, if $x \mapsto g(x)$ is a polynomial in x we write $p_{g(x)}$ to denote the function $x \mapsto g(x)$. We can simplify further using integration by parts. It is clear that f vanishes at infinity faster than any polynomial, and the expression for f' as a difference of gaussians shows this is also true of every derivative of f. Thus, we find straightforwardly

$$\begin{aligned}
\int_{\mathbb{R}} p_{x^3} f f' &= -\frac{3}{2}\|p_x f\|_{L^2}^2, \\
\int_{\mathbb{R}} p_x f' f'' &= -\frac{1}{2}\|f'\|_{L^2}^2, \\
\int_{\mathbb{R}} f f'' &= -\|f'\|_{L^2}^2, \\
\int_{\mathbb{R}} p_x f f' &= -\frac{1}{2}\|f\|_{L^2}^2, \\
\int_{\mathbb{R}} p_{x^2} f f'' &= \|f\|_{L^2}^2 - \|p_x f'\|_{L^2}^2.
\end{aligned}$$

Applying these identities, we simplify the previous expression to

$$\begin{aligned}
\int_{\mathbb{R}^2} \left(\left\langle \nabla_{\boldsymbol{x}} \mathcal{C}(s,t), \begin{bmatrix} -t \\ s \end{bmatrix} \right\rangle_{\ell^2} \right)^2 \mathrm{d}s\,\mathrm{d}t &= 6\|p_x f'\|_{L^2}^2 \left(\|p_x f'\|_{L^2}^2 - \|f\|_{L^2}^2\right) + \frac{5}{2}\|f\|_{L^2}^4 \\
&\quad + 2\|f''\|_{L^2}^2 \|p_{x^2} f\|_{L^2}^2 - 10\|f'\|_{L^2}^2 \|p_x f\|_{L^2}^2 \\
&\leq 6\|p_x f'\|_{L^2}^4 + \frac{5}{2}\|f\|_{L^2}^4 + 2\|f''\|_{L^2}^2 \|p_{x^2} f\|_{L^2}^2
\end{aligned}$$

We can estimate the integrals involving f using Jensen's inequality. In particular, notice that

$$\begin{aligned}(f(s))^2 &= \left(\int_{-\alpha}^{\alpha} \varphi_{\sigma^2}(s-x)\,\mathrm{d}x\right)^2 \\ &= (2\alpha)^2\left(\frac{1}{2\alpha}\int_{-\alpha}^{\alpha} \varphi_{\sigma^2}(s-x)\,\mathrm{d}x\right)^2 \\ &\leq 2\alpha\int_{-\alpha}^{\alpha} \varphi_{\sigma^2}(s-x)^2\,\mathrm{d}x,\end{aligned}$$

by Jensen's inequality for the convex function $x \mapsto x^2$. Since

$$\varphi_{\sigma^2}(s-x)^2 = \frac{1}{2\sqrt{\pi\sigma^2}}\varphi_{\sigma^2/2}(s-x),$$

we obtain

$$(f(s))^2 \leq \frac{\alpha}{\sigma\sqrt{\pi}}\int_{-\alpha}^{\alpha} \varphi_{\sigma^2/2}(s-x)\,\mathrm{d}x,$$

which is a scaled version of f with the variance of the gaussian smoothing halved. From here, it follows by Fubini's theorem and standard (non-centered) gaussian moment calculations

$$\begin{aligned}\|f\|_{L^2}^2 &\leq \frac{2\alpha^2}{\sigma\sqrt{\pi}}; \\ \|p_{x^2}f\|_{L^2}^2 &\leq \frac{\alpha}{\sigma\sqrt{\pi}}\int_{-\alpha}^{\alpha}\int_{\mathbb{R}} s^4\varphi_{\sigma^2/2}(s-x)\,\mathrm{d}s\,\mathrm{d}x \\ &= \frac{\alpha}{4\sigma\sqrt{\pi}}\int_{-\alpha}^{\alpha}\left(3\sigma^4 + 12\sigma^2x^2 + 4x^4\right)\mathrm{d}x \\ &= \frac{\alpha^2\left(15\sigma^4 + 20\sigma^2\alpha^2 + 4\alpha^4\right)}{10\sigma\sqrt{\pi}}.\end{aligned}$$

The remaining terms are gaussian integrals, and can be calculated easily. We evaluate

$$\begin{aligned}\|p_x f'\|_{L^2}^2 &= \frac{1}{2\sigma\sqrt{\pi}}\left(2\alpha^2 + \sigma^2\left(1 - e^{-\frac{\alpha^2}{\sigma^2}}\right)\right); \\ \|f''\|_{L^2}^2 &= \frac{1}{2\sigma^5\sqrt{\pi}}\left(2\alpha^2 e^{-\frac{\alpha^2}{\sigma^2}} + \sigma^2\left(1 - e^{-\frac{\alpha^2}{\sigma^2}}\right)\right).\end{aligned}$$

We can simplify these expressions further: applying the inequality $1 - x \leq e^{-x}$ gives

$$\begin{aligned}\|p_x f'\|_{L^2}^2 &\leq \frac{3\alpha^2}{2\sigma\sqrt{\pi}}; \\ \|f''\|_{L^2}^2 &\leq \frac{3\alpha^2}{2\sigma^5\sqrt{\pi}}.\end{aligned}$$

Combining, we thus get

$$\int_{\mathbb{R}^2}\left(\left\langle \nabla_{\boldsymbol{x}}\mathcal{C}(s,t), \begin{bmatrix}-t\\ s\end{bmatrix}\right\rangle_{\ell^2}\right)^2 \mathrm{d}s\,\mathrm{d}t \leq \frac{28\alpha^4}{\pi\sigma^2} + \frac{3\alpha^6(20\sigma^2 + 4\alpha^2)}{10\pi\sigma^6}.$$

We can obtain improved estimates when σ is small: writing

$$\|f''\|_{L^2}^2 = \frac{1}{2\sigma^3\sqrt{\pi}}\left(\frac{2\alpha^2}{\sigma^2}e^{-\frac{\alpha^2}{\sigma^2}} + 1 - e^{-\frac{\alpha^2}{\sigma^2}}\right),$$

evidently

$$\frac{2\alpha^2}{\sigma^2}e^{-\frac{\alpha^2}{\sigma^2}} + 1 - e^{-\frac{\alpha^2}{\sigma^2}} \leq 1 + \frac{2}{e},$$

and the RHS is no larger than 2; hence

$$\|f''\|_{L^2}^2 \leq \frac{1}{\sigma^3\sqrt{\pi}}.$$

Combining in this case gives the estimate (together with $\alpha^2 \leq 1$)

$$\int_{\mathbb{R}^2} \left(\left\langle \nabla_{\boldsymbol{x}} \mathcal{C}(s,t), \begin{bmatrix} -t \\ s \end{bmatrix} \right\rangle_{\ell^2} \right)^2 \mathrm{d}s\,\mathrm{d}t \leq \frac{3}{\pi} + \frac{55}{\pi\sigma^2} + \frac{4}{5\pi\sigma^4}.$$

□

F.3. Auxiliary Results

Lemma F.15. *Let $f : [-1,+1] \to \mathbb{R}$ be a L-Lipschitz function, and let $\pi_1(G)$ be the projection of the rectangular grid G onto its first coordinate. One has*

$$\left| \frac{2}{n} \sum_{i \in \pi_1(G)} f(i) - \int_{[-1,1]} f(t)\,\mathrm{d}t \right| \leq \frac{2L}{n}.$$

Proof. Define

$$\delta_i = -1 + i\frac{2}{n-1}, \quad i = 0, 1, \ldots, n-1,$$

so that

$$\int_{[-1,1]} f(t)\,\mathrm{d}t = \int_{\delta_0}^{\delta_0+\frac{2}{n}} f(t)\,\mathrm{d}t + \int_{\delta_{n-1}-\frac{2}{n}}^{\delta_{n-1}} f(t)\,\mathrm{d}t + \sum_{i=1}^{n-2} \int_{\delta_i-\frac{1}{n}}^{\delta_i+\frac{1}{n}} f(t)\,\mathrm{d}t.$$

This is a 'midpoint' estimate of the integral, given the boundary. Since

$$\sum_{i \in \pi_1(G)} f(i) = \sum_{i=0}^{n-1} f(\delta_i),$$

we obtain from the triangle inequality and the Lipschitz property of f

$$\begin{aligned}
\left| \frac{2}{n} \sum_{i \in \pi_1(G)} f(i) - \int_{-1}^{1} f(t)\,\mathrm{d}t \right| &\leq \int_{\delta_0}^{\delta_0+\frac{2}{n}} |f(\delta_0) - f(t)|\,\mathrm{d}t + \int_{\delta_{n-1}-\frac{2}{n}}^{\delta_{n-1}} |f(\delta_{n-1}) - f(t)|\,\mathrm{d}t + \sum_{i=1}^{n-2} \int_{\delta_i-\frac{1}{n}}^{\delta_i+\frac{1}{n}} |f(\delta_i) - f(t)|\,\mathrm{d}t \\
&\leq L \left(\int_{\delta_0}^{\delta_0+\frac{2}{n}} (t - \delta_0)\,\mathrm{d}t + \int_{\delta_{n-1}-\frac{2}{n}}^{\delta_{n-1}} (\delta_{n-1} - t)\,\mathrm{d}t + \sum_{i=1}^{n-2} \int_{\delta_i-\frac{1}{n}}^{\delta_i+\frac{1}{n}} |t - \delta_i|\,\mathrm{d}t \right) \\
&= L \left(\frac{4}{n^2} + \sum_{i=1}^{n-2} \frac{1}{n^2} \right) \\
&\leq \frac{2L}{n},
\end{aligned}$$

where the last estimate holds if $n \geq 2$.

□

Lemma F.16. *Let $\boldsymbol{U}, \boldsymbol{V} \in \mathbb{R}^{m\times n}$, and let $\boldsymbol{D} \in \mathbb{R}^{n\times n}$ be a diagonal matrix. Let $|||\cdot|||$ be any unitarily invariant matrix norm. Then one has*

$$|||\boldsymbol{U}\boldsymbol{D}\boldsymbol{V}^*||| \leq \frac{1}{2}\left(|||\boldsymbol{D}|^{1/2}\boldsymbol{U}^*\boldsymbol{U}|\boldsymbol{D}|^{1/2}||| + |||\boldsymbol{D}|^{1/2}\boldsymbol{V}^*\boldsymbol{V}|\boldsymbol{D}|^{1/2}||| \right),$$

where $|\boldsymbol{A}| = (\boldsymbol{A}^*\boldsymbol{A})^{1/2}$ denotes the positive part of a matrix, and the matrix norms in this expression are to be interpreted in terms of the 'dilation norm' of the larger size matrix norm.[5]

Proof. We apply a slight modification of a matrix arithmetic-geometric mean inequality. There exists a diagonal matrix $\boldsymbol{S} \in \mathbb{R}^{n\times n}$ with diagonal entries either 1 or $\sqrt{-1}$ such that $\boldsymbol{SDS} = |\boldsymbol{D}|$. Then $\boldsymbol{S}^*\boldsymbol{S} = \boldsymbol{I}$, so $\boldsymbol{S}$ is unitary, and by [7, Corollary IX.4.4],

$$\begin{aligned}|||\boldsymbol{UDV}^*||| = |||\boldsymbol{US}|\boldsymbol{D}|\boldsymbol{SV}^*||| &= |||(\boldsymbol{US}|\boldsymbol{D}|^{1/2})(\boldsymbol{VS}^*|\boldsymbol{D}|^{1/2})^*||| \\ &\leq \frac{1}{2}\left(|||\boldsymbol{D}|^{1/2}\boldsymbol{S}^*\boldsymbol{U}^*\boldsymbol{US}|\boldsymbol{D}|^{1/2} + |\boldsymbol{D}|^{1/2}\boldsymbol{SV}^*\boldsymbol{VS}^*|\boldsymbol{D}|^{1/2}|||\right).\end{aligned}$$

Now apply the triangle inequality and use the fact that diagonal matrices commute and that $|||\cdot|||$ is unitarily invariant to establish the claim. □

Lemma F.17. *For $\sigma^2 > 0$, let $\varphi_{\sigma^2}(t) = 1/\sqrt{2\pi\sigma^2}\exp(-\frac{1}{2\sigma^2}t^2)$ denote the one-dimensional standard gaussian, and let $m_{\sigma^2} = \varphi_{\sigma^2}^{\otimes 2}$. Let $f, g \in L^1(\mathbb{R}^2) \cap L^2(\mathbb{R}^2)$, and let $\bar{G}$ denote the infinite extension of the image sampling grid G defined in (F.6):*

$$\bar{G} = \left\{ \left(1 + \frac{2k}{n-1}, 1 + \frac{2l}{n-1}\right) \,\middle|\, (k,l) \in \mathbb{Z}^2 \right\}$$

(notice that $G \subset \bar{G}$). Let $\ell^2(\bar{G})$ denote the space of square-summable sequences defined on $\bar{G}$. Then it holds

$$\left| \left(\frac{n-1}{2}\right)^2 \langle m_{\sigma^2} * f, m_{\sigma^2} * g\rangle_{L^2(\mathbb{R}^2)} - \langle m_{\sigma^2} * f, m_{\sigma^2} * g\rangle_{\ell^2(\bar{G})} \right| \leq \frac{\|f\|_{L^1}\|g\|_{L^1}}{(2\pi\sigma^2)^2}\left(1 + \frac{(n-1)\sigma}{\sqrt{2}}\right).$$

Proof. We will rely on machinery from the theory of tempered distributions throughout the proof, following notation and results contained in [4, Ch. I, §3]. Let $\mathscr{S} \subset L^2(\mathbb{R}^2)$ denote the class of real-valued Schwartz functions (a dense subset of $L^2(\mathbb{R}^2)$). For concision, write $a = m_{\sigma^2} * f$ and $b = m_{\sigma^2} * g$. Then because f, g are in L^1, a, b are in $\mathscr{S}$. Let $\delta_{\boldsymbol{x}}$ denote the "Dirac distribution" at $\boldsymbol{x} \in \mathbb{R}^2$, the tempered distribution defined by $\delta_{\boldsymbol{x}}(h) = h(\boldsymbol{x})$ for every $h \in \mathscr{S}$. Let Δ_n denote the "Dirac comb" for the grid $\bar{G}$, the tempered distribution defined by

$$\Delta_n = \sum_{(i,j)\in\bar{G}} \delta_{(i,j)}.$$

Notice that when n is odd, we have $\bar{G} = (2/(n-1))\mathbb{Z}^2$, and when n is even we have $\bar{G} = 1/(n-1) + (2/(n-1))\mathbb{Z}^2$. Then from the definition of the product, convolution, and Fourier transform of tempered distributions, we have

$$\begin{aligned}\langle a, b\rangle_{\ell^2(\bar{G})} &= \Delta_n(ab) \\ &= (\Delta_n a)(b) \\ &= (\hat{\Delta}_n * \hat{a})\hat{}\,(b) \qquad \text{(F.91)} \\ &= (\hat{\Delta}_n * \hat{a})(\hat{b}) \\ &= \hat{\Delta}_n(\tilde{\hat{a}} * \hat{b}), \qquad \text{(F.92)}\end{aligned}$$

where $*$ additionally denotes convolution of a tempered distribution with a Schwartz function, for a Schwartz function or a tempered distribution $\hat{\psi}$ denotes its Fourier transform, and for a Schwartz function $\tilde{g}$ denotes its reversal $\tilde{g}(\boldsymbol{x}) = g(-\boldsymbol{x})$. Above, (F.91) applies the convolution formula for tempered distributions (c.f. [4, Proof of Ch. I, Thm. 3.18]), and the remaining

[5] That is, if $n > m$, $|||\cdot|||$ is the matrix norm on $n \times n$ matrices, and if $\boldsymbol{A} \in \mathbb{R}^{m\times m}$

$$|||\boldsymbol{A}||| = |||\begin{bmatrix}\boldsymbol{A} & \boldsymbol{0}\\ \boldsymbol{0} & \boldsymbol{0}\end{bmatrix}|||,$$

and likewise if $m > n$. Compare [7, Exercise IV.2.15].

manipulations are unraveling definitions. The tempered distribution $\hat{\varphi}$ is defined by the relation $\hat{\varphi}(h) = \varphi(\hat{h})$ for all $h \in \mathscr{S}$; so we have for the Dirac comb and for any $h \in \mathscr{S}$

$$\begin{aligned}\hat{\Delta}_n(h) = \sum_{(i,j)\in\bar{G}} \delta_{(i,j)}(\hat{h}) &= \sum_{(k,l)\in\mathbb{Z}^2} \int_{\mathbb{R}^2} h(\boldsymbol{x}) e^{-\mathrm{i}2\pi\left(\left\langle \frac{2\boldsymbol{x}}{n-1},(k,l)\right\rangle + \mathbb{1}_{\text{n even}}\left\langle \frac{\boldsymbol{x}}{n-1},(1,1)\right\rangle\right)} \,\mathrm{d}\boldsymbol{x} \\ &= \left(\frac{n-1}{2}\right)^2 \sum_{(k,l)\in\mathbb{Z}^2} \int_{\mathbb{R}^2} h\left(\frac{n-1}{2}\boldsymbol{x}\right) e^{-\mathrm{i}\pi\langle \boldsymbol{x}, \mathbb{1}_{\text{n even}}(1,1)\rangle} e^{-\mathrm{i}2\pi\langle \boldsymbol{x},(k,l)\rangle} \,\mathrm{d}\boldsymbol{x} \\ &= \left(\frac{n-1}{2}\right)^2 \sum_{(k,l)\in\mathbb{Z}^2} \left(h_{\frac{n-1}{2}} \cdot e^{-\mathrm{i}\pi\langle\,\cdot\,,(1,1)\mathbb{1}_{\text{n even}}\rangle}\right)\hat{}\,(k,l),\end{aligned}$$

where in the final line $h_{(n-1)/2}$ denotes the dilation of h (as in the previous line). Now, because $h \in \mathscr{S}$, it holds that $\bar{h}(\boldsymbol{x}) = h_{\frac{n-1}{2}}(\boldsymbol{x}) e^{-\mathrm{i}\pi\langle \boldsymbol{x},(1,1)\rangle}$ satisfies $\bar{h} \in \mathscr{S}$, because the complex exponential function is infinitely differentiable with uniformly bounded derivatives on $\mathbb{R}^2$. We can thus apply the Poisson summation formula [4, Ch. VII, Cor. 2.6] to obtain from the previous

$$\begin{aligned}\hat{\Delta}_n(h) &= \left(\frac{n-1}{2}\right)^2 \sum_{(k,l)\in\mathbb{Z}^2} \left(h_{\frac{n-1}{2}} \cdot e^{-\mathrm{i}\pi\langle\,\cdot\,,(1,1)\mathbb{1}_{\text{n even}}\rangle}\right)(k,l) \\ &= \left(\frac{n-1}{2}\right)^2 \sum_{(k,l)\in\mathbb{Z}^2} e^{-\mathrm{i}\pi k\mathbb{1}_{\text{n even}}} e^{-\mathrm{i}\pi l\mathbb{1}_{\text{n even}}} h_{\frac{n-1}{2}}(k,l).\end{aligned}$$

This shows that $\hat{\Delta}_n$ is equal to a modulated Dirac comb on a rescaled grid. Continuing from (F.92), we therefore have

$$\begin{aligned}\langle a,b\rangle_{\ell^2(\bar{G})} &= \left(\frac{n-1}{2}\right)^2 \left[\sum_{(k,l)\in\mathbb{Z}^2} e^{-\mathrm{i}\pi k\mathbb{1}_{\text{n even}}} e^{-\mathrm{i}\pi l\mathbb{1}_{\text{n even}}} \int_{\mathbb{R}^2} \hat{a}(\boldsymbol{\xi})\hat{b}(\boldsymbol{\xi} + \tfrac{n-1}{2}(k,l)) \,\mathrm{d}\boldsymbol{\xi}\right] \\ &= \left(\frac{n-1}{2}\right)^2 \left[\int_{\mathbb{R}^2} \hat{a}(\boldsymbol{\xi})\hat{b}(\boldsymbol{\xi}) \,\mathrm{d}\boldsymbol{\xi} + \sum_{\substack{(k,l)\in\mathbb{Z}^2 \\ (k,l)\neq\mathbf{0}}} e^{-\mathrm{i}\pi k\mathbb{1}_{\text{n even}}} e^{-\mathrm{i}\pi l\mathbb{1}_{\text{n even}}} \int_{\mathbb{R}^2} \hat{a}(\boldsymbol{\xi})\hat{b}(\boldsymbol{\xi} + \tfrac{n-1}{2}(k,l)) \,\mathrm{d}\boldsymbol{\xi}\right],\end{aligned}$$

where we applied a change of variables to simplify the convolution integrals to cross-correlations. Now, by Parseval's theorem on Schwartz functions, we have

$$\langle a,b\rangle_{\ell^2(\bar{G})} = \left(\frac{n-1}{2}\right)^2 \langle a,b\rangle_{L^2(\mathbb{R}^2)} + \left(\frac{n-1}{2}\right)^2 \sum_{\substack{(k,l)\in\mathbb{Z}^2 \\ (k,l)\neq\mathbf{0}}} e^{-\mathrm{i}\pi k\mathbb{1}_{\text{n even}}} e^{-\mathrm{i}\pi l\mathbb{1}_{\text{n even}}} \int_{\mathbb{R}^2} \hat{a}(\boldsymbol{\xi})\hat{b}(\boldsymbol{\xi} + \tfrac{n-1}{2}(k,l)) \,\mathrm{d}\boldsymbol{\xi}, \tag{F.93}$$

so our task is to bound the residual in the previous expression. We have $\hat{a} = (\varphi_{\sigma^2}^{\otimes 2})\hat{}\, \hat{f}$ by the convolution formula for L^2 functions (and similarly for $\hat{b}$), and the Fourier transform of a gaussian is another gaussian, suitably scaled ([4, Theorem 1.13]):

$$(\varphi_{\sigma^2}^{\otimes 2})\hat{}\,(\boldsymbol{\xi}) = e^{-2\pi^2\sigma^2\|\boldsymbol{\xi}\|_2^2} = \frac{1}{2\pi\sigma^2}\varphi_{1/(2\pi\sigma)^2}^{\otimes 2}.$$

Because $f, g \in L^1(\mathbb{R}^2)$, we have $\|\hat{f}\|_{L^\infty} \leq \|f\|_{L^1}$ and $\|\hat{g}\|_{L^\infty} \leq \|g\|_{L^1}$. For the residual term in (F.93), we thus have the

estimate

$$\left(\frac{n-1}{2}\right)^2 \left| \sum_{\substack{(k,l)\in\mathbb{Z}^2 \\ (k,l)\neq \mathbf{0}}} e^{-\mathrm{i}\pi k \mathbb{1}_{\text{n even}}} e^{-\mathrm{i}\pi l \mathbb{1}_{\text{n even}}} \int_{\mathbb{R}^2} \hat{g}(\boldsymbol{\xi}) \hat{g}(\boldsymbol{\xi} + \tfrac{n-1}{2}(k,l)) \,\mathrm{d}\boldsymbol{\xi} \right|$$

$$\leq \|f\|_{L^1} \|g\|_{L^1} \left(\frac{n-1}{4\pi\sigma^2}\right)^2 \sum_{\substack{(k,l)\in\mathbb{Z}^2 \\ (k,l)\neq \mathbf{0}}} \int_{\mathbb{R}^2} \varphi^{\otimes 2}_{1/(2\pi\sigma)^2}(\boldsymbol{\xi}) \varphi^{\otimes 2}_{1/(2\pi\sigma)^2}(\boldsymbol{\xi} + \tfrac{n-1}{2}(k,l)) \,\mathrm{d}\boldsymbol{\xi}, \tag{F.94}$$

where we applied the triangle inequality. The integral in the previous expression is a convolution integral; as is well-known, the convolution of two gaussians is another gaussian, with mean equal to the sum of the means of the factors and variance equal to the sum of the variances. In particular, we have (using reflection symmetry of the gaussian function)

$$\int_{\mathbb{R}^2} \varphi^{\otimes 2}_{1/(2\pi\sigma)^2}(\boldsymbol{\xi}) \varphi^{\otimes 2}_{1/(2\pi\sigma)^2}(\boldsymbol{\xi} + \tfrac{n-1}{2}(k,l)) \,\mathrm{d}\boldsymbol{\xi} = \varphi^{\otimes 2}_{2/(2\pi\sigma)^2}\left(\tfrac{n-1}{2}(k,l)\right).$$

Because the gaussian function factors across components of its argument, we have

$$\sum_{\substack{(k,l)\in\mathbb{Z}^2 \\ (k,l)\neq \mathbf{0}}} \varphi^{\otimes 2}_{2/(2\pi\sigma)^2}\left(\tfrac{n-1}{2}(k,l)\right) = \left(\sum_{k\in\mathbb{Z}} \varphi_{2/(2\pi\sigma)^2}\left(\tfrac{n-1}{2}k\right)\right)^2 - \left(\varphi_{2/(2\pi\sigma)^2}(0)\right)^2.$$

Let $\gamma^2 = 2/(2\pi\sigma^2)$. We have

$$\begin{aligned}
\varphi_{2/(2\pi\sigma)^2}\left(\tfrac{n-1}{2}k\right) &= \frac{1}{\sqrt{2\pi\gamma^2}} e^{-\frac{1}{2\gamma^2}((n-1)/2)^2 k^2} \\
&= \frac{2}{n-1} \frac{1}{\sqrt{2\pi\bar{\gamma}^2}} e^{-\frac{1}{2\bar{\gamma}^2}k^2} \\
&= \frac{2}{n-1} \varphi_{\bar{\gamma}^2}(k),
\end{aligned}$$

where we have defined $\bar{\gamma}^2 = \gamma^2/((n-1)/2)^2$. Estimating the sum with the integral test estimate gives

$$\begin{aligned}
\sum_{k\in\mathbb{Z}} \varphi_{\bar{\gamma}^2}(k) &\leq 2\left(\varphi_{\bar{\gamma}^2}(0) + \int_0^\infty \varphi_{\bar{\gamma}^2}(\xi) \,\mathrm{d}\xi\right) - \varphi_{\bar{\gamma}^2}(0) \\
&= 1 + \varphi_{\bar{\gamma}^2}(0),
\end{aligned}$$

so in particular

$$\begin{aligned}
\left(\sum_{k\in\mathbb{Z}} \varphi_{2/(2\pi\sigma)^2}\left(\tfrac{n-1}{2}k\right)\right)^2 - \left(\varphi_{2/(2\pi\sigma)^2}(0)\right)^2 &\leq \left(\frac{2}{n-1} + \varphi_{\gamma^2}(0)\right)^2 - \left(\varphi_{\gamma^2}(0)\right)^2 \\
&= \left(\frac{2}{n-1}\right)^2 + \frac{4\varphi_{\gamma^2}(0)}{n-1}.
\end{aligned}$$

With this, (F.94) can be bounded as

$$\left(\frac{n-1}{2}\right)^2 \left| \sum_{\substack{(k,l)\in\mathbb{Z}^2 \\ (k,l)\neq \mathbf{0}}} e^{-\mathrm{i}\pi k \mathbb{1}_{\text{n even}}} e^{-\mathrm{i}\pi l \mathbb{1}_{\text{n even}}} \int_{\mathbb{R}^2} \hat{g}(\boldsymbol{\xi}) \hat{g}(\boldsymbol{\xi} + \tfrac{n-1}{2}(k,l)) \,\mathrm{d}\boldsymbol{\xi} \right| \leq \|f\|_{L^1} \|g\|_{L^1} \left(\frac{1}{2\pi\sigma^2}\right)^2 \left(1 + \frac{(n-1)\sigma}{\sqrt{2}}\right),$$

which implies the claim.

□

Lemma F.18. *Let $f, g \in L^1(\mathbb{R}) \cap L^2(\mathbb{R})$, and for $\sigma^2 > 0$, let $\varphi_{\sigma^2}(t) = 1/\sqrt{2\pi\sigma^2}\exp(-\frac{1}{2\sigma^2}t^2)$ denote the one-dimensional standard gaussian. Then one has*

$$|\langle f, g\rangle_{L^2} - \langle \varphi_{\sigma^2} * f, \varphi_{\sigma^2} * g\rangle_{L^2}| \leq \sigma^2 \|f'\|_{L^2(\mathbb{R}^2)}\|g'\|_{L^2(\mathbb{R}^2)}.$$

Proof. One calculates with Plancherel's theorem and the convolution theorem for the Fourier transform

$$\begin{aligned}
|\langle f, g\rangle_{L^2} - \langle \varphi_{\sigma^2} * f, \varphi_{\sigma^2} * g\rangle_{L^2}| &= \left|\langle \hat{\varphi}_{\sigma^2}\hat{f}, \hat{\varphi}_{\sigma^2}\hat{g}\rangle - \langle \hat{f}, \hat{g}\rangle\right| \\
&= \left|\langle (\hat{\varphi}_{\sigma^2})^2\hat{f}, \hat{g}\rangle - \langle \hat{f}, \hat{g}\rangle\right| \\
&= \left|\langle ((\hat{\varphi}_{\sigma^2})^2 - 1)\hat{f}, \hat{g}\rangle\right| \\
&= \left|\langle \sqrt{1-(\hat{\varphi}_{\sigma^2})^2}\hat{f}, \sqrt{1-(\hat{\varphi}_{\sigma^2})^2}\hat{g}\rangle\right|,
\end{aligned}$$

where we use that the Fourier transform of a gaussian is another gaussian (and in particular, is positive and bounded by 1):

$$\hat{\varphi}_{\sigma^2}(\xi) = e^{-2\pi^2\sigma^2\xi^2}.$$

We thus have, by the triangle inequality,

$$\begin{aligned}
\left|\langle \sqrt{1-(\hat{\varphi}_{\sigma^2})^2}\hat{f}, \sqrt{1-(\hat{\varphi}_{\sigma^2})^2}\hat{g}\rangle\right| &\leq \int_{\mathbb{R}^2} |\hat{f}\hat{g}|(\xi)(1 - e^{-4\pi^2\sigma^2\xi^2})\,\mathrm{d}\xi \\
&\leq \sigma^2 \int_{\mathbb{R}^2} |\mathrm{i}2\pi\xi\hat{f}(\xi)||\mathrm{i}2\pi\xi\hat{g}(\xi)|\,\mathrm{d}\xi
\end{aligned}$$

using $1 - e^{-x} \leq x$ in the second line. By [4, Theorem §I, 1.8], we have $\mathrm{i}2\pi\xi\hat{f}(\xi) = (f')\hat{}(\xi)$, whence by the Schwarz inequality and Parseval's theorem

$$\sigma^2 \int_{\mathbb{R}^2} |\mathrm{i}2\pi\xi\hat{f}(\xi)||\mathrm{i}2\pi\xi\hat{g}(\xi)|\,\mathrm{d}\xi \leq \sigma^2\|f'\|_{L^2(\mathbb{R}^2)}\|g'\|_{L^2(\mathbb{R}^2)}.$$

□

Lemma F.19. *For $X \in L^2(\mathbb{R}^2)$, consider the rank-one factorization objective*

$$\min_{u \in L^2(\mathbb{R})} \frac{1}{2}\|X - uu^*\|^2_{L^2(\mathbb{R}^2)}.$$

Suppose that there exists a nonzero $v \in L^2(\mathbb{R})$ such that $v^*\mathcal{T}_X v \geq 0$, where $\mathcal{T}_X : L^2(\mathbb{R}) \to L^2(\mathbb{R})$ denotes the integral operator $u \mapsto \int_{\mathbb{R}} X(\,\cdot\,, t)u(t)\,\mathrm{d}t$ associated to X. Then this optimization problem is equivalent to the constrained problem

$$\max_{\|u\|_{L^2(\mathbb{R})}=1} u^*\left(\mathcal{T}_X + \mathcal{T}_X^*\right)u;$$

precisely, if u is an optimal solution to the second problem, then $(\frac{1}{2}u^*(\mathcal{T}_X + \mathcal{T}_X^*)u)uu^*$ is an optimal solution to the first problem. Here, $\mathcal{T}_X^*$ is the adjoint of $\mathcal{T}_X$.

Proof. The problem

$$\min_{u \in L^2(\mathbb{R})} \frac{1}{2}\|X - uu^*\|^2_{L^2(\mathbb{R}^2)}.$$

is equivalent to the problem

$$\min_{\|u\|_{L^2(\mathbb{R})}=1,\, c\geq 0} \frac{1}{2}\|X - cuu^*\|^2_{L^2(\mathbb{R}^2)}.$$

Expanding the square, the objective in this latter problem satisfies

$$\|X - cuu^*\|^2_{L^2(\mathbb{R}^2)} = \|X\|^2_{L^2(\mathbb{R}^2)} - 2c\langle X, uu^*\rangle_{L^2(\mathbb{R}^2)} + c^2,$$

since u is constrained to be unit norm. By elementary calculus, the minimization over c in this problem can be calculated in closed form; we find that the optimal c is equal to $\langle X, uu^* \rangle_{L^2(\mathbb{R}^2)} = u^* \mathcal{T}_X u = \frac{1}{2} u^* (\mathcal{T}_X + \mathcal{T}_X^*) u$, where $\mathcal{T}_X^* u = \int_{\mathbb{R}} X(s, \cdot) u(s) \, \mathrm{d}s$ is the adjoint of $\mathcal{T}_X$. Hence, the original problem is equivalent to the problem

$$\min_{\|u\|_{L^2(\mathbb{R})}=1, u^*(\mathcal{T}_X + \mathcal{T}_X^*)u \geq 0} \frac{1}{2} \left\| X - (\tfrac{1}{2} u^* (\mathcal{T}_X + \mathcal{T}_X^*) u) uu^* \right\|_{L^2(\mathbb{R}^2)}^2.$$

Expanding the square as before, this objective satisfies

$$\frac{1}{2} \left\| X - (\tfrac{1}{2} u^* (\mathcal{T}_X + \mathcal{T}_X^*) u) uu^* \right\|_{L^2(\mathbb{R}^2)}^2 = \|X\|_{L^2(\mathbb{R}^2)}^2 - (\tfrac{1}{2} u^* (\mathcal{T}_X + \mathcal{T}_X^*) u)^2$$

at any point where $u^*(\mathcal{T}_X + \mathcal{T}_X^*)u \geq 0$; otherwise, the objective equals $\|X\|_{L^2(\mathbb{R}^2)}^2$. Now, if for every nonzero u we have $u^*(\mathcal{T}_X + \mathcal{T}_X^*)u < 0$, then evidently the only optimal solution to the problem is $u = 0$. If for some nonzero u we have $u^*(\mathcal{T}_X + \mathcal{T}_X^*)u \geq 0$, then the previous expression shows that the problem is equivalent to

$$\max_{\|u\|_{L^2(\mathbb{R})}=1} u^* (\mathcal{T}_X + \mathcal{T}_X^*) u,$$

which is feasible.

□

F.4. Background on Image Resampling

We give a precise definition of the vector field representation underlying (F.2) in the discrete setting (F.29). For the template image $\boldsymbol{X}_\natural \in \mathbb{R}^{m \times n}$, $\boldsymbol{X}_\natural \circ \boldsymbol{\tau}_\nu$ denotes image resampling:

$$\boldsymbol{X}_\natural \circ \boldsymbol{\tau}_\nu = \sum_{(k,l) \in G} (\boldsymbol{X}_\natural)_{kl} \phi\left(\tfrac{n-1}{2}\left(\boldsymbol{\tau}_\nu^0 - k\mathbf{1}\mathbf{1}^*\right)\right) \odot \phi\left(\tfrac{n-1}{2}\left(\boldsymbol{\tau}_\nu^1 - l\mathbf{1}\mathbf{1}^*\right)\right). \tag{F.95}$$

Here, $\phi : \mathbb{R} \to \mathbb{R}$ is the interpolation kernel; it is applied elementwise, and is independent of the image content and resolution. Typical choices for this kernel in practice are the bilinear interpolation kernel (which is continuous, but not continuously differentiable; we adopt it in our experiments) and the cubic convolution interpolation kernel [5] (which is continuously differentiable, with an absolutely continuous derivative). Both of these kernels are compactly supported, which allows (F.95) to be computed with cost proportional to the image size. The transformation field $\boldsymbol{\tau}_\nu \in \mathbb{R}^{m \times n \times 2}$ is defined as

$$\boldsymbol{\tau}_\nu^0 = \cos\nu \left(\frac{2}{n-1}\boldsymbol{n} - \mathbf{1}\right)\mathbf{1}^* + \sin\nu \, \mathbf{1}\left(\frac{2}{n-1}\boldsymbol{n} - \mathbf{1}\right)^* \tag{F.96}$$

$$\boldsymbol{\tau}_\nu^1 = -\sin\nu \left(\frac{2}{n-1}\boldsymbol{n} - \mathbf{1}\right)\mathbf{1}^* + \cos\nu \, \mathbf{1}\left(\frac{2}{n-1}\boldsymbol{n} - \mathbf{1}\right)^* \tag{F.97}$$

where $\boldsymbol{n} = [0, 1, \ldots, n-1]$ (c.f. [30, §A.1] and (F.6)). Note that this definition ensures that the resampled image $\boldsymbol{X}_\natural \circ \boldsymbol{\tau}_\nu$ corresponds to a rotation of the image content by an angle of ν (with the usual "counterclockwise" positive orientation): in particular,

$$(\boldsymbol{\tau}_\nu)_{ij} = \begin{bmatrix} \cos\nu & -\sin\nu \\ \sin\nu & \cos\nu \end{bmatrix}^* \begin{bmatrix} i \\ j \end{bmatrix}$$

for $(i, j) \in G$ defined in (F.6).

References

[1] Carl Eckart and Gale Young, "The approximation of one matrix by another of lower rank," *Psychometrika*, vol. 1, no. 3, pp. 211–218, Sep. 1936. 8.

[2] L Mirsky, "SYMMETRIC GAUGE FUNCTIONS AND UNITARILY INVARIANT NORMS," *The Quarterly Journal of Mathematics*, vol. 11, no. 1, pp. 50–59, Jan. 1960. 8.

[3] Chandler Davis and W M Kahan, "The rotation of eigenvectors by a perturbation. III," *SIAM journal on numerical analysis*, vol. 7, no. 1, pp. 1–46, Mar. 1970. 27.

[4] Elias M Stein and Guido Weiss, *Introduction to Fourier Analysis on Euclidean Spaces*, en. Princeton University Press, 1971. 15, 57, 58, 60.

[5] R Keys, "Cubic convolution interpolation for digital image processing," *IEEE transactions on acoustics, speech, and signal processing*, vol. 29, no. 6, pp. 1153–1160, Dec. 1981. 61.

[6] J Kuczyński and H Woźniakowski, "Estimating the largest eigenvalue by the power and lanczos algorithms with a random start," *SIAM Journal on Matrix Analysis and Applications*, vol. 13, no. 4, pp. 1094–1122, Oct. 1992. 30.

[7] Rajendra Bhatia, *Matrix Analysis*. Springer, New York, NY, 1997. 8, 57.

[8] Jor-Ting Chan, Chi-Kwong Li, and Charlies Tu, "A class of unitarily invariant norms on b(h)," en, *Proceedings of the American Mathematical Society. American Mathematical Society*, vol. 129, no. 4, pp. 1065–1076, Oct. 2000. 7.

[9] Martin Lefébure and Laurent D Cohen, "Image registration, optical flow and local rigidity," *Journal of mathematical imaging and vision*, vol. 14, no. 2, pp. 131–147, Mar. 2001. 28.

[10] Yurii Nesterov, *Introductory Lectures on Convex Optimization: A Basic Course* (Applied Optimization), 1st ed. Springer US, 2004. 29.

[11] Haim Brezis, *Functional Analysis, Sobolev Spaces and Partial Differential Equations*. Springer, New York, NY, 2011. 30.

[12] Christopher Heil, *A Basis Theory Primer: Expanded Edition*. Birkhäuser Boston, 2011. 6, 8, 25.

[13] Diederik P Kingma and Jimmy Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, 2014. 1.

[14] Benjamin D Haeffele and Rene Vidal, "Global optimality in tensor factorization, deep learning, and beyond," Jun. 2015. arXiv: 1506.07540 [cs.NA]. 20.

[15] Rong Ge, Chi Jin, and Yi Zheng, "No spurious local minima in nonconvex low rank problems: A unified geometric analysis," Apr. 2017. arXiv: 1704.00708 [cs.LG]. 20.

[16] Yuanzhi Li, Tengyu Ma, and Hongyang Zhang, "Algorithmic regularization in over-parameterized matrix sensing and neural networks with quadratic activations," Dec. 2017. arXiv: 1712.09203 [cs.LG]. 20.

[17] Gary Bécigneul and Octavian-Eugen Ganea, "Riemannian adaptive optimization methods," *arXiv preprint arXiv:1810.00760*, 2018. 1.

[18] Yuejie Chi, Yue M Lu, and Yuxin Chen, "Nonconvex optimization meets Low-Rank matrix factorization: An overview," Sep. 2018. arXiv: 1809.09573 [cs.LG]. 19, 20, 29.

[19] Yu Bai, Qijia Jiang, and Ju Sun, "Subgradient descent learns orthogonal dictionaries," in *International Conference on Learning Representations*, 2019. 29.

[20] Dar Gilboa, Sam Buchanan, and John Wright, "Efficient dictionary learning with gradient descent," in *Proceedings of the 36th International Conference on Machine Learning*, Kamalika Chaudhuri and Ruslan Salakhutdinov, Eds., ser. Proceedings of Machine Learning Research, vol. 97, Long Beach, California, USA: PMLR, 2019, pp. 2252–2259. 29.

[21] Tero Karras, Samuli Laine, Miika Aittala, Janne Hellsten, Jaakko Lehtinen, and Timo Aila, "Analyzing and improving the image quality of stylegan," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2020, pp. 8110–8119. 5.

[22] Yuqian Zhang, Qing Qu, and John Wright, "From symmetry to geometry: Tractable nonconvex problems," Jul. 2020. arXiv: 2007.06753 [cs.LG]. 20, 22, 29.

[23] Zhimin Zhang, Jinpan Fang, Jaduo Lin, Shancheng Zhao, Fengjun Xiao, and Jinming Wen, "Improved upper bound on the complementary error function," en, *Electronics letters*, vol. 56, no. 13, pp. 663–665, Jun. 2020. 45.

[24] Jonathan T. Barron, Ben Mildenhall, Dor Verbin, Pratul P. Srinivasan, and Peter Hedman, "Mip-nerf 360: Unbounded anti-aliased neural radiance fields," *arXiv*, 2021. 1.

[25] Eric R. Chan, Connor Z. Lin, Matthew A. Chan, Koki Nagano, Boxiao Pan, Shalini De Mello, Orazio Gallo, Leonidas Guibas, Jonathan Tremblay, Sameh Khamis, Tero Karras, and Gordon Wetzstein, "Efficient geometry-aware 3D generative adversarial networks," in *arXiv*, 2021. 4, 5.

[26] Chen-Hsuan Lin, Wei-Chiu Ma, Antonio Torralba, and Simon Lucey, "Barf: Bundle-adjusting neural radiance fields," in *IEEE International Conference on Computer Vision (ICCV)*, 2021. 3.

[27] Keunhong Park, Utkarsh Sinha, Jonathan T Barron, Sofien Bouaziz, Dan B Goldman, Steven M Seitz, and Ricardo Martin-Brualla, "Nerfies: Deformable neural radiance fields," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2021, pp. 5865–5874. 3.

[28] Robin Rombach, Andreas Blattmann, Dominik Lorenz, Patrick Esser, and Björn Ommer, *High-resolution image synthesis with latent diffusion models*, 2021. arXiv: 2112.10752 [cs.CV]. 5.

[29] Dominik Stöger and Mahdi Soltanolkotabi, "Small random initialization is akin to spectral learning: Optimization and generalization guarantees for overparameterized low-rank matrix reconstruction," Jun. 2021. arXiv: 2106.15013 [cs.LG]. 19, 20, 29.

[30] Sam Buchanan, Jingkai Yan, Ellie Haber, and John Wright, "Resource-Efficient invariant networks: Exponential gains by unrolled optimization," Mar. 2022. arXiv: 2203.05006 [cs.CV]. 61.

[31] Anpei Chen, Zexiang Xu, Andreas Geiger, Jingyi Yu, and Hao Su, "Tensorf: Tensorial radiance fields," in *European Conference on Computer Vision (ECCV)*, 2022. 4, 5.

[32] Thomas Müller, Alex Evans, Christoph Schied, and Alexander Keller, "Instant neural graphics primitives with a multiresolution hash encoding," *arXiv:2201.05989*, Jan. 2022. 5.

[33] Tengfei Wang, Bo Zhang, Ting Zhang, Shuyang Gu, Jianmin Bao, Tadas Baltrusaitis, Jingjing Shen, Dong Chen, Fang Wen, Qifeng Chen, and Baining Guo, *Rodin: A generative model for sculpting 3d digital avatars using diffusion*, arXiv, Dec. 2022. 4, 5.

[34] Anpei Chen, Zexiang Xu, Xinyue Wei, Siyu Tang, Hao Su, and Andreas Geiger, "Factor fields: A unified framework for neural fields and beyond," *arXiv preprint arXiv:2302.01226*, 2023. 1.

[35] Sara Fridovich-Keil, Giacomo Meanti, Frederik Warburg, Benjamin Recht, and Angjoo Kanazawa, "K-planes: Explicit radiance fields in space, time, and appearance," *arXiv preprint arXiv:2301.10241*, 2023. 5.

[36] Matthew Tancik, Ethan Weber, Evonne Ng, Ruilong Li, Brent Yi, Justin Kerr, Terrance Wang, Alexander Kristoffersen, Jake Austin, Kamyar Salahi, Abhik Ahuja, David McAllister, and Angjoo Kanazawa, "Nerfstudio: A modular framework for neural radiance field development," *arXiv preprint arXiv:2302.04264*, 2023. 1.

[37] Rachel Ward and Tamara G Kolda, "Convergence of alternating gradient descent for matrix factorization," May 2023. arXiv: 2305.06927 [cs.LG]. 20.

[38] Xingyu Xu, Yandi Shen, Yuejie Chi, and Cong Ma, "The power of preconditioning in overparameterized Low-Rank matrix sensing," Feb. 2023. arXiv: 2302.01186 [cs.LG]. 20.