Linear Frequency Principle

§1. neural tangent kernel (NTK) & frequency principle (FP)

- 2-layer NN: $f(x;\theta) = f_\theta(x) = \sum_{k=1}^{m} a_k \sigma(w_k \cdot x + b_k)$

  parameter: $\theta = \text{vec}\{a_k, w_k, b_k\}_{k=1}^{m}$

  unknown: D, f

  sample : $S = \{(x_i, y_i)\}_{i=1}^{n}$, $\{x_i\}_{i=1}^{n} \overset{\text{i.i.d.}}{\sim} D(\Omega)$, $y_i = f(x_i)$

  $\Omega \subset \mathbb{R}^d$

  empirical risk $R_S(\theta) = \frac{1}{2}\sum_{i=1}^{n} (f_\theta(x_i) - y_i)^2$

  gradient descent (GD) :
  $$\begin{cases} \dot{\theta} = -\nabla_\theta R_S(\theta) \\ \theta(0) = \theta_0 \end{cases} \qquad (1)$$

- linearization of NN

$$f^{lin}(x,\theta) = f^{lin}(x,\theta_0) + \nabla_\theta f^{lin}(x,\theta_0)\cdot(\theta - \theta_0)$$

$$R_S^{lin}(\theta) = \frac{1}{2}\sum_{i=1}^{n} (f^{lin}(x_i,\theta) - y_i)^2$$

GD :
$$\begin{cases} \dot{\theta} = -\nabla_\theta R_S^{lin}(\theta) \\ \theta(0) = \theta_0 \end{cases}$$

still denote as

$$f(x,\theta) = f(x,\theta_0) + \nabla_\theta f(x,\theta_0)\cdot(\theta - \theta_0)$$

$$R_S(\theta) = \frac{1}{2}\sum_{i=1}^{n} (f(x_i,\theta) - y_i)^2$$

GD :
$$\begin{cases} \dot{\theta} = -\nabla_\theta R_S(\theta) \\ \theta(0) = \theta_0 \end{cases} \qquad (2)$$

Remark : $\nabla_\theta f(x,\theta) = \nabla_\theta f(x,\theta_0)$

- neural tangent kernel (NTK)

$$(2) \Rightarrow \frac{d}{dt} f(x,\theta) = \nabla_\theta f(x,\theta) \cdot \dot{\theta}$$
$$= -\nabla_\theta f(x,\theta) \cdot \nabla_\theta R_S(\theta)$$
$$= -\nabla_\theta f(x,\theta) \cdot \sum_{i=1}^{n} \nabla_\theta f(x_i,\theta)\left(f(x_i,\theta) - y_i\right)$$
$$= -\sum_{i=1}^{n} K_m(x, x_i)\left(f(x_i,\theta) - y_i\right)$$

NTK: $K_m(x, x') = \nabla_\theta f(x,\theta) \cdot \nabla_\theta f(x',\theta)$

$= \nabla_\theta f(x,\theta_0) \cdot \nabla_\theta f(x',\theta_0)$

$x, x' \in \Omega$

only depends on θ_0

linear dynamics with NTK:

$$\frac{d}{dt}\left(f(x,\theta(t)) - f(x)\right) = -\sum_{i=1}^{n} K_m(x, x_i)\left(f(x_i,\theta(t)) - f(x_i)\right) \qquad (3)$$

Let $u(x,t) = f(x,\theta(t)) - f(x)$ ($u(x,t)$ depends on θ)

$\rho(x) = \sum_{i=1}^{n} \delta(x_i)$

$u_\rho(x,t) = u(x,t)\,\rho(x)$

$$\Rightarrow \quad \frac{d}{dt} u(x,t) = -\int_{\mathbb{R}^d} K_m(x,x')\, u_\rho(x',t)\, dx' \qquad (4)$$

- NTK for 2-layer NN

$$f_\theta(x) = \frac{1}{\sqrt{m}} \sum_{k=1}^{m} a_k\, \sigma(w_k \cdot x + b_k) = \frac{1}{\sqrt{m}} \sum_{k=1}^{m} \sigma^*(x, q_k)$$

generalized coordinates $q_k = (a_k, w_k, b_k)$

$$K_m(x,x') = \frac{1}{m}\nabla_\theta f(x,\theta_0)\cdot\nabla_\theta f(x',\theta_0) = \frac{1}{m}\sum_{k=1}^{m}\nabla_{q_k}\sigma^*(x,q_k)\cdot\nabla_{q_k}\sigma^*(x',q_k)$$

$$= \frac{1}{m}\sum_{k=1}^{m}\Big(\nabla_{a_k}\sigma^*(x,q_k^0)\cdot\nabla_{a_k}\sigma^*(x',q_k^0)$$
$$+\nabla_{w_k}\sigma^*(x,q_k^0)\cdot\nabla_{w_k}\sigma^*(x',q_k^0)$$
$$+\nabla_{b_k}\sigma^*(x,q_k^0)\cdot\nabla_{b_k}\sigma^*(x',q_k^0)\Big)$$

$$= \frac{1}{m}\sum_{k=1}^{m}\Big(\sigma(w_k^0 x+b_k^0)\,\sigma(w_k^0\cdot x'+b_k^0)$$
$$+a_k^0\sigma'(w_k^0 x+b_k^0)\,a_k^0\sigma'(w_k^0\cdot x'+b_k^0)\,(x\cdot x')$$
$$+a_k^0\sigma'(w_k^0\cdot x+b_k^0)\,a_k^0\sigma'(w_k^0\cdot x'+b_k^0)\Big)$$

as $m\to\infty$, $K_m(x,x') = K(x,x') + O\left(\frac{1}{\sqrt{m}}\right)$

$$K(x,x') := \mathbb{E}_q \nabla_q\sigma^*(x,q)\cdot\nabla_q\sigma^*(x',q)$$
$$= \mathbb{E}_q\Big(\sigma(w\cdot x+b)\,\sigma(w\cdot x'+b) + a^2\sigma'(w\cdot x+b)\,\sigma'(w\cdot x'+b)\,(x\cdot x')$$
$$+ a^2\sigma'(w\cdot x+b)\,\sigma'(w\cdot x'+b)\Big)$$

$q=(a,w,b)$, usually
$a\sim N(0,\sigma_a^2)$
$w\sim N(0,\sigma_w^2 I_d)$
$b\sim N(0,\sigma_b^2)$

- frequency principle (FP)

first observed by experiments, later some rigorous results for general DNN (or even continuous parametric models) are available.

FP phenomenon: DNNs often fit target functions from low to high frequencies in (GD) training dynamics

Question: for GD of DNN, can we write down its dependence on different frequencies?

Answer:

This can be done for the GD of a very wide ($m \to +\infty$) linearized 2-layer NN.

- informal derivation of LFP (linear frequency principle)

recall $\frac{d}{dt} u(x) = -\int_{\mathbb{R}^d} k(x,x') u_p(x') dx' + O(\frac{1}{\sqrt{m}})$

drop lower order term $O(\frac{1}{\sqrt{m}})$

regard $u(\cdot)$ as a column vector, $k(\cdot,\cdot)$ as a square matrix

$$\Rightarrow \text{"} \frac{d}{dt} u = -K u_p \text{"}$$

regard the Fourier transform as a matrix

$$\mathcal{F} : u \mapsto \mathcal{F}[u] = \hat{u}$$

$$\text{"} u \mapsto F u \text{"}$$

$$\mathcal{F}^{-1} : u \mapsto \mathcal{F}^{-1}[u]$$

$$\text{"} u \mapsto F^{-1} u \text{"}$$

$$\Rightarrow \text{"} \frac{d}{dt} F u = -F K u_p$$

$$= -(F K F^{-1}) F u_p$$

$$\frac{d}{dt} \hat{u} = -\hat{K} \hat{u}_p \text{"}$$

$$\text{"} \hat{K} := F K F^{-1} \text{"}$$

goal: calculate $\hat{K}$ rigorously.

§2 derivation of LFP dynamics

- Fourier transform on $\mathbb{R}^d$

  for $g \in L^1(\mathbb{R}^d)$

  $$\mathcal{F}[g](\xi) := \mathcal{F}_{x\to\xi}[g](\xi) := \int_{\mathbb{R}^d} g(x) e^{-2\pi i x\cdot\xi} dx$$

  inverse Fourier transform

  $$\mathcal{F}^{-1}[g](x) := \mathcal{F}^{-1}_{\xi\to x}[g](x) := \int_{\mathbb{R}^d} g(\xi) e^{2\pi i x\cdot\xi} d\xi$$

  Recall: σ = ReLU, tanh, ...

  not smooth, not in L^2 ...

  L^1/L^2 theory of Fourier transform is not suitable

  distribution / generalized function is required

  Schwartz space $\mathcal{S}(\mathbb{R}^d)$

  space of tempered distributions $\mathcal{S}'(\mathbb{R}^d)$

  $$\psi \in \mathcal{S}'(\mathbb{R}^d),\ \phi \in \mathcal{S}(\mathbb{R}^d),\ \langle \psi, \phi \rangle := \langle \psi, \phi \rangle_{\mathcal{S}'(\mathbb{R}^d), \mathcal{S}(\mathbb{R}^d)} := \psi(\phi)$$

  Fourier transform of $\psi \in \mathcal{S}'(\mathbb{R}^d)$

  $$\langle \mathcal{F}[\psi], \phi \rangle = \langle \psi, \mathcal{F}[\phi] \rangle \quad \forall \phi \in \mathcal{S}(\mathbb{R}^d)$$

- Prop. 1. $\langle \partial_t \mathcal{F}[u], \phi \rangle = \langle \mathcal{L}[\mathcal{F}[u_p]], \phi \rangle + O\left(\frac{1}{\sqrt{m}}\right)$

  $\mathcal{L}[\cdot]$ is called LFP operator

  $$\mathcal{L}[\mathcal{F}[u_p]] = -\int_{\mathbb{R}^d} \hat{K}(\xi, \xi') \mathcal{F}[u_p](\xi') d\xi'$$

$$\hat{K}(\xi, \xi') := \mathbb{E}_g \hat{K}_g(\xi, \xi')$$
$$:= \mathbb{E}_g \mathcal{F}_{x\to\xi}[\nabla_g \sigma^*(x, g)] \cdot \overline{\mathcal{F}_{x'\to\xi'}[\nabla_g \sigma^*(x', g)]}$$

Proof: Recall

$$\partial_t u(x,t) = -\int_{\mathbb{R}^d} K(x, x')\, u_p(x', t)\, dx' + O\left(\tfrac{1}{\sqrt{m}}\right)$$
$$\approx -\int_{\mathbb{R}^d} K(x, x')\, u_p(x', t)\, dx'.$$

use "$\approx$" to keep leading order

$\forall \phi \in \mathcal{S}(\mathbb{R}^d)$

$$\langle \mathcal{F}[\partial_t u], \phi \rangle = \langle \partial_t u, \mathcal{F}[\phi] \rangle$$

$$= \int_{\mathbb{R}^d} \partial_t u(x,t) \int_{\mathbb{R}^d} \phi(\xi)\, e^{-i2\pi x\cdot\xi}\, d\xi\, dx \quad (\because \partial_t u \text{ locally integrable})$$

$$\approx -\int_{\mathbb{R}^d} \int_{\mathbb{R}^d} K(x, x')\, u_p(x', t)\, dx' \int_{\mathbb{R}^d} \phi(\xi)\, e^{-i2\pi x\cdot\xi}\, d\xi\, dx$$

$$= -\int_{\mathbb{R}^{3d}} K(x, x')\, u_p(x', t)\, dx'\, \phi(\xi)\, e^{-i2\pi x\cdot\xi}\, d\xi\, dx$$

$$= -\int_{\mathbb{R}^{3d}} \mathbb{E}_g \nabla_g \sigma^*(x, g) \cdot \nabla_g \sigma^*(x', g)\, u_p(x', t)\, dx'\, \phi(\xi)\, e^{-i2\pi x\cdot\xi}\, d\xi\, dx$$

$$= -\mathbb{E}_g \int_{\mathbb{R}^d} \nabla_g \sigma^*(x', g)\, u_p(x', t)\, dx' \cdot \int_{\mathbb{R}^{2d}} \nabla_g \sigma^*(x, g)\, e^{-i2\pi x\cdot\xi} \phi(\xi)\, d\xi\, dx$$

$$= -\mathbb{E}_g \int_{\mathbb{R}^d} \nabla_g \sigma^*(x', g)\, u_p(x', t)\, dx' \cdot \langle \mathcal{F}_{x\to\cdot}[\nabla_g \sigma^*(x, g)](\cdot), \phi(\cdot) \rangle$$

Since $\int_{\mathbb{R}^d} \nabla_q \sigma^*(x', q) u_p(x') dx'$

$$= \int_{\mathbb{R}^d} \overline{\mathcal{F}_{x' \to \xi'}[\nabla_q \sigma^*(x', q)](\xi')} \mathcal{F}_{x' \to \xi'}[u_p](\xi') d\xi'$$

$$\langle \partial_t \mathcal{F}[u], \phi \rangle$$

$$\approx - \mathbb{E}_q \int_{\mathbb{R}^d} \overline{\mathcal{F}_{x' \to \xi'}[\nabla_q \sigma^*(x', q)](\xi')} \mathcal{F}_{x' \to \xi'}[u_p](\xi') d\xi' \cdot \langle \mathcal{F}_{x \to \cdot}[\nabla_q \sigma^*(x, q)](\cdot), \phi(\cdot) \rangle$$

$$= - \int_{\mathbb{R}^{2d}} \mathbb{E}_q \overline{\mathcal{F}_{x' \to \xi'}[\nabla_q \sigma^*(x', q)](\xi')} \cdot \mathcal{F}_{x \to \xi}[\nabla_q \sigma^*(x, q)](\xi) \mathcal{F}_{x' \to \xi'}[u_p](\xi') d\xi' \phi(\xi) d\xi$$

$$= - \int_{\mathbb{R}^{2d}} \hat{K}(\xi, \xi') \mathcal{F}[u_p](\xi') d\xi' \phi(\xi) d\xi$$

$$= \langle \mathcal{L}[\mathcal{F}[u_p]], \phi \rangle$$

\#

§3 explicit expression for LFP dynamics with ReLU activation

- Def: delta-like function $\delta_w : \mathcal{S}(\mathbb{R}^d) \to \mathbb{R}$

$$\langle \delta_w, \phi \rangle = \int_{\mathbb{R}} \phi(yw) dy$$

- Lem 1: $\forall$ nonzero $w \in \mathbb{R}^d$, let $\hat{w} = \frac{w}{\|w\|}$

then $$\frac{1}{\|w\|^d} \delta_{\hat{w}} \left(\frac{x}{\|w\|} \right) = \delta_w(x)$$

Proof: $\forall \phi \in \mathcal{S}(\mathbb{R}^d)$

$$\left\langle \frac{1}{\|w\|^d} \delta_{\hat{w}} \left(\frac{\cdot}{\|w\|} \right), \phi(\cdot) \right\rangle = \langle \delta_{\hat{w}}(\cdot), \phi(\|w\| \cdot) \rangle$$
$$= \int_{\mathbb{R}} \phi(\|w\| y \hat{w}) dy$$
$$= \int_{\mathbb{R}} \phi(yw) dy$$

$= \langle \delta_w(\cdot), \phi(\cdot) \rangle$ #

8

· Lem 2: $\forall$ unit vector $\nu \in \mathbb{R}^d$ $\forall$ nonzero $w \in \mathbb{R}^d$

let $\hat{w} := \frac{w}{\|w\|}$ $\forall g \in S'(\mathbb{R}^d)$

then in the sense of distribution

(a) $\mathcal{F}_{x\to\xi}[g(\nu\cdot x)](\xi) = \delta_\nu(\xi)\, \mathcal{F}[g](\xi\cdot\nu)$

(b) $\mathcal{F}_{x\to\xi}[g(w\cdot x+b)](\xi) = \delta_w(\xi)\, \mathcal{F}[g]\left(\frac{\xi\cdot\hat{w}}{\|w\|}\right) e^{2\pi i \frac{b}{\|w\|}\xi\cdot\hat{w}}$

(c) $\mathcal{F}_{x\to\xi}[x g(w\cdot x+b)](\xi) = \frac{i}{2\pi}\nabla_\xi\left[\delta_w(\xi)\mathcal{F}[g]\left(\frac{\xi\cdot\hat{w}}{\|w\|}\right) e^{2\pi i \frac{b}{\|w\|}\xi\cdot\hat{w}}\right]$

Proof: $\forall \phi \in S(\mathbb{R}^d)$

(a) $\langle \mathcal{F}_{x\to\cdot}[g(\nu\cdot x)](\cdot), \phi(\cdot)\rangle_{S'(\mathbb{R}^d), S(\mathbb{R}^d)}$

$= \langle g(\nu^T\cdot), \mathcal{F}_{x\to\cdot}[\phi(x)](\cdot)\rangle_{S'(\mathbb{R}^d), S(\mathbb{R}^d)}$

$= \langle g(\cdot), \mathcal{F}_{y\to\cdot}[\phi(y\nu)](\cdot)\rangle_{S'(\mathbb{R}), S(\mathbb{R})}$

$= \langle \mathcal{F}_{y\to\cdot}[g(y)](\cdot), \phi(\cdot\,\nu)\rangle_{S'(\mathbb{R}), S(\mathbb{R})}$

$= \langle \mathcal{F}[g](\cdot\,\nu^T\nu), \phi(\cdot\,\nu)\rangle_{S'(\mathbb{R}), S(\mathbb{R})}$

$= \langle \delta_\nu(\cdot)\mathcal{F}[g](\cdot^T\nu), \phi(\cdot)\rangle_{S'(\mathbb{R}^d), S(\mathbb{R}^d)}$

(b) $\mathcal{F}_{x\to\xi}[g(\hat{w}\cdot x)](\xi) = \delta_{\hat{w}}(\xi)\mathcal{F}[g](\xi\cdot\hat{w})$

note that

$$\mathcal{F}_{x\to\xi}[g(x-x_0)](\xi) = \mathcal{F}_{x\to\xi}[g](\xi)e^{-2\pi i x_0\cdot\xi}$$

then

$$\mathcal{F}_{x\to\xi}[g(\hat{w}\cdot x+b)](\xi) = \mathcal{F}_{x\to\xi}[g(\hat{w}\cdot(x+b\hat{w}))](\xi)$$

$$= \delta_w(\xi)\, \mathcal{F}[g](\xi\cdot\hat{w})\, e^{2\pi i b \hat{w}\cdot\xi}$$ 9

$\therefore$ $\mathcal{F}_{x\to\xi}[g(w\cdot x+b)](\xi)$

$$= \mathcal{F}_{x\to\xi}[g(\hat{w}\cdot\|w\|x+b)](\xi)$$

$$= \frac{1}{\|w\|^d}\mathcal{F}_{x\to\xi}[g(\hat{w}\cdot x+b)]\left(\frac{\xi}{\|w\|}\right)$$

$$= \frac{1}{\|w\|^d}\,\delta_{\hat{w}}\left(\frac{\xi}{\|w\|}\right)\mathcal{F}[g]\left(\frac{\xi\cdot\hat{w}}{\|w\|}\right)e^{2\pi i \frac{b}{\|w\|}\hat{w}\cdot\xi}$$

$$= \delta_w(\xi)\,\mathcal{F}[g]\left(\frac{\xi\cdot\hat{w}}{\|w\|}\right)e^{2\pi i \frac{b}{\|w\|}\hat{w}\cdot\xi}$$

(c) (b) & $\forall\ \tilde{g}\in S'(\mathbb{R}^d)$

$$\mathcal{F}_{x\to\xi}[x\,\tilde{g}(x)](\xi) = \frac{i}{2\pi}\nabla_\xi[\mathcal{F}[\tilde{g}](\xi)]$$

$\Rightarrow$ (c) #

- apply the lemma with

$$g_1(w\cdot x+b) = \begin{pmatrix}\partial_a[a\sigma(w\cdot x+b)]\\ \partial_b[a\sigma(w\cdot x+b)]\end{pmatrix}$$

and $x\, g_2(w\cdot x+b) = \nabla_w[a\,\sigma(w\cdot x+b)]$

- Thm $\langle\partial_t\mathcal{F}[u],\phi\rangle = \langle\mathcal{L}[\mathcal{F}[u_p]],\phi\rangle + O(m^{-1/2}) + O(\sigma_b^{-3})$

$\mathcal{L}[\mathcal{F}[u_p]](\xi)$

$$= -\frac{1}{(2\pi)^{5/2}\sigma_b\|\xi\|^{d+1}}\mathbb{E}_{a,r}\left[\frac{1}{r}\mathcal{F}[g_1]\left(\frac{\|\xi\|}{r}\right)\cdot\mathcal{F}[g_1]\left(-\frac{\|\xi\|}{r}\right)\right]\mathcal{F}[u_p](\xi)$$

$$+\frac{1}{(2\pi)^{5/2}\sigma_b}\nabla\cdot\left\{\mathbb{E}_{a,r}\left[\frac{1}{r\|\xi\|^{d+1}}\mathcal{F}[g_2]\left(\frac{\|\xi\|}{r}\right)\mathcal{F}[g_2]\left(-\frac{\|\xi\|}{r}\right)\right]\nabla\mathcal{F}[u_p](\xi)\right\}$$

where $r = \|w\|$

Remark: if $\sigma = \mathrm{ReLU}$, then

$$\mathcal{L}[\mathcal{F}[u_\rho]](\xi)$$
$$= -\frac{1}{(2\pi)^{5/2}\sigma_b}\mathbb{E}_{a,r}\left[\frac{r^3}{16\pi^4\|\xi\|^{d+3}} + \frac{a^2 r}{4\pi^2\|\xi\|^{d+1}}\right]\mathcal{F}[u_\rho](\xi)$$
$$+ \frac{1}{(2\pi)^{5/2}\sigma_b}\nabla\cdot\left\{\mathbb{E}_{a,r}\left[\frac{a^2 r}{4\pi^2\|\xi\|^{d+1}}\right]\nabla\mathcal{F}[u_\rho](\xi)\right\}$$

Proof. $\hat{k} = \mathbb{E}_g \hat{k}_g$, $\hat{k}_g = \hat{k}_{a,b} + \hat{k}_w$

$$\hat{k}_w(\xi,\xi') = \mathcal{F}_{x\to\xi}[x\, g_2(w\cdot x+b)](\xi)\cdot\overline{\mathcal{F}_{x'\to\xi'}[x' g_2(w\cdot x'+b)](\xi')}$$
$$\hat{k}_{a,b}(\xi,\xi') = \mathcal{F}_{x\to\xi}[g_1(w\cdot x+b)](\xi)\cdot\overline{\mathcal{F}_{x'\to\xi'}[g_1(w\cdot x'+b)](\xi')}$$
$$= \delta_w(\xi)\delta_w(\xi')\mathcal{F}[g_1]\left(\frac{\xi\cdot\hat{w}}{\|w\|}\right)\cdot\overline{\mathcal{F}[g_1]\left(\frac{\xi'\cdot\hat{w}}{\|w\|}\right)}\, e^{2\pi i b(\xi-\xi')\cdot\frac{\hat{w}}{\|w\|}}$$

$$\langle\phi,\psi\rangle_{k_{a,b}} := \int_{\mathbb{R}^d}\phi(\xi)\int_{\mathbb{R}^d}\hat{k}_{a,b}(\xi,\xi')\psi(\xi')\,d\xi'\,d\xi$$
$$= \int_{\mathbb{R}\times\mathbb{R}}\phi(\eta w)\psi(\eta' w)\mathcal{F}[g_1](\eta)\cdot\overline{\mathcal{F}[g_1](\eta')}\, e^{2\pi i b(\eta-\eta')}\,d\eta\,d\eta'$$

Recall that $b\sim\mathcal{N}(0,\sigma_b^2)$

$$\rho_b(b) = \frac{1}{\sqrt{2\pi}\sigma_b}e^{-\frac{b^2}{2\sigma_b^2}}$$
$$\mathcal{F}[\rho_b](\eta) = e^{-2\pi^2\sigma_b^2\eta^2}$$
$$\mathbb{E}_b\left(e^{2\pi i b(\eta-\eta')}\right) = \int_{\mathbb{R}}\frac{1}{\sqrt{2\pi}\sigma_b}e^{-b^2/2\sigma_b^2}e^{2\pi i b(\eta-\eta')}\,db$$
$$= \mathcal{F}[\rho_b](-(\eta-\eta'))$$
$$= e^{-2\pi^2\sigma_b^2(\eta-\eta')^2}$$

$$\mathbb{E}_b \langle \phi,\psi\rangle_{K_{a,b}} = \int_{\mathbb{R}\times\mathbb{R}} \phi(\eta w)\psi(\eta' w)\mathcal{F}[g_1](\eta)\cdot\overline{\mathcal{F}[g_1](\eta')}\,\mathbb{E}_b\left[e^{2\pi i b(\eta-\eta')}\right]d\eta\, d\eta'$$

$$= \int_{\mathbb{R}\times\mathbb{R}} \phi(\eta w)\psi(\eta' w)\mathcal{F}[g_1](\eta)\cdot\overline{\mathcal{F}[g_1](\eta')}\, e^{-2\pi^2\sigma_b^2(\eta-\eta')^2}\, d\eta\, d\eta'$$

$$= \int_{\mathbb{R}} \phi(\eta w)\mathcal{F}[g_1](\eta)\cdot\left[\int_{\mathbb{R}} \psi(\eta' w)\overline{\mathcal{F}[g_1](\eta')}\, e^{-2\pi^2\sigma_b^2(\eta-\eta')^2}\, d\eta'\right]d\eta$$

Laplace method
$$= \int_{\mathbb{R}} \phi(\eta w)\mathcal{F}[g_1](\eta)\left[\psi(\eta w)\overline{\mathcal{F}[g_1](\eta)}\,\frac{1}{\sqrt{2\pi}\,\sigma_b} + O(\sigma_b^{-3})\right]d\eta$$

$$= \frac{1}{\sqrt{2\pi}\,\sigma_b}\int_{\mathbb{R}} \phi(\eta w)\psi(\eta w)\mathcal{F}[g_1](\eta)\cdot\overline{\mathcal{F}[g_1](\eta)}\,d\eta + O(\sigma_b^{-3})$$

Keep leading order

$$\mathbb{E}_{w,b} \langle \phi,\psi\rangle_{K_{a,b}} \approx \mathbb{E}_w \frac{1}{\sqrt{2\pi}\,\sigma_b}\int_{\mathbb{R}} \phi(\eta w)\psi(\eta w)\mathcal{F}[g_1](\eta)\cdot\overline{\mathcal{F}[g_1](\eta)}\,d\eta$$

$$= \int_{\mathbb{R}^{d+1}} \frac{1}{\sqrt{2\pi}\,\sigma_b}\phi(\eta w)\psi(\eta w)\mathcal{F}[g_1](\eta)\cdot\overline{\mathcal{F}[g_1](\eta)}\,\rho_w(w)\,dw\,d\eta$$

assume ρ_w is radially symmetric, i.e., $\rho_w(w) = \rho_w(r) = \rho_w(\|w\|)$

$$\rho_r(r) := \int_{\mathbb{S}^{d-1}} \rho_w(r)\, r^{d-1}\, d\hat{w} = 4\pi^2 \rho_w(r)\, r^{d-1}$$

change of variables

$$\begin{cases} \zeta = \eta w \\ r = \|w\| \end{cases}$$

$$\Rightarrow \quad \det\left(\frac{\partial(w,\eta)}{\partial(\zeta,r)}\right) = -\frac{r^{d-1}}{r\|\zeta\|^{d-1}} \quad \text{(Jacobian determinant)}$$

$\therefore$ $\mathbb{E}_{w,b} \langle \phi,\psi\rangle_{K_{a,b}}$

$$\approx \int_{\mathbb{R}^{d+1}} \frac{1}{\sqrt{2\pi}\,\sigma_b}\phi(\eta w)\psi(\eta w)\mathcal{F}[g_1](\eta)\cdot\overline{\mathcal{F}[g_1](\eta)}\,\rho_w(r)\,dw\,d\eta$$

$$= \int_{\mathbb{R}^d \times \mathbb{R}^+} \frac{1}{\sqrt{2\pi}\sigma_b} \phi(\zeta)\psi(\zeta) \mathcal{F}[g_1]\left(\frac{\|\zeta\|}{r}\right) \cdot \overline{\mathcal{F}[g_1]\left(\frac{\|\zeta\|}{r}\right)} \frac{r^{d+1}}{r\|\zeta\|^{d+1}} \rho_w(r) d\zeta dr$$

$$= \int_{\mathbb{R}^d \times \mathbb{R}^+} \frac{1}{\sqrt{2\pi}\sigma_b} \phi(\zeta)\psi(\zeta) \mathcal{F}[g_1]\left(\frac{\|\zeta\|}{r}\right) \cdot \overline{\mathcal{F}[g_1]\left(\frac{\|\zeta\|}{r}\right)} \frac{r^{d+1}}{r\|\zeta\|^{d+1}} \frac{\rho_r(r)}{4\pi^2} d\zeta dr$$

$$= \frac{1}{(2\pi)^{d/2}\sigma_b} \int_{\mathbb{R}^d} \phi(\zeta) \int_{\mathbb{R}^+} \left[\frac{1}{r\|\zeta\|^{d+1}} \mathcal{F}[g_1]\left(\frac{\|\zeta\|}{r}\right) \cdot \overline{\mathcal{F}[g_1]\left(\frac{\|\zeta\|}{r}\right)}\right] \psi(\zeta) \rho_r(r) \, dr \, d\zeta$$

taking $\psi = \mathcal{F}[u_p]$, we have

$$\mathcal{L}_{a,b}[\mathcal{F}[u_p]] = -\frac{1}{(2\pi)^{d/2}\sigma_b\|\zeta\|^{d+1}} \mathbb{E}_{a,r}\left[\frac{1}{r}\mathcal{F}[g_1]\left(\frac{\|\zeta\|}{r}\right) \cdot \overline{\mathcal{F}[g_1]\left(\frac{\|\zeta\|}{r}\right)}\right] \mathcal{F}[u_p](\zeta)$$

$$= -\frac{1}{(2\pi)^{d/2}\sigma_b\|\zeta\|^{d+1}} \mathbb{E}_{a,r}\left[\frac{1}{r}\mathcal{F}[g_1]\left(\frac{\|\zeta\|}{r}\right) \cdot \mathcal{F}[g_1]\left(-\frac{\|\zeta\|}{r}\right)\right] \mathcal{F}[u_p](\zeta)$$

similarly, from R_w we can derive

$\mathcal{L}_w[\mathcal{F}[u_p]]$

$$= \frac{1}{(2\pi)^{d/2}\sigma_b} \nabla \cdot \left\{\mathbb{E}_{a,r}\left[\frac{1}{r\|\zeta\|^{d+1}} \mathcal{F}[g_2]\left(\frac{\|\zeta\|}{r}\right) \cdot \mathcal{F}[g_2]\left(-\frac{\|\zeta\|}{r}\right)\right] \nabla \mathcal{F}[u_p](\zeta)\right\}$$

#