Compositional Reproducing Kernel

Ohad Asor

January 10, 2016

Recalling basic RKHS (Reproducing Kernel Hilbert Space) theory deals with bivariate functions that reproduce point-evaluation by inner-product, or formally:

$$\psi = \langle \psi, k(\psi, \cdot) \rangle \tag{1}$$

The classical illustrations are over the Hilbert space $L_2[\mathcal{I}]$ of all square integrable functions over some (possibly infinite) interval $\mathcal{I}$. Suppose $\mathcal{H} \subset L_2$ is a Hilbert subspace and there exists a differential operator $D(\psi) : \mathcal{H} \to L_2$ and a function k such that

$$\psi(x) = \int_{\mathcal{I}} D(\psi)(x)\, k(x,t)\, dt \tag{2}$$

for all $\psi \in \mathcal{H}$. If the mapping $\psi \to \langle \psi, k(\psi, \cdot) \rangle$ is continuous, then $\mathcal{H}$ is called a RKHS. We can then prove many extremely useful and interesting properties of such spaces and the kernel function, as has been done for over a century (argubaly even more than two centuries at the scope of the Heat equation). To mention at least one reproducing kernel, suppose $\mathcal{H}$ is the space of all functions f such that their Fourier transform $\hat{f}$ satisfy $\hat{f}(z) = 0$ for all $|z| \leq 1$. Then the reproducing kernel together with the reproducing property can be written as

$$\psi(x) = \frac{1}{\pi} \int_{\mathbb{R}} \psi(x) \frac{\sin[\pi(x-t)]}{x-t} dt \tag{3}$$

The reproducing property is recovered via an inner product. On this note we present methods to reproduce a function in rather compositional manner, than lacks the linear properties of inner products.

Given f to approximate and assume it is holomorphic on a disc of radius r, consider by Cauchy's integral theorem

$$I = f(z_0) = \frac{1}{2\pi i} \int_{|z|=r} \frac{f(z)}{z - z_0} dz = \frac{1}{2\pi i} \int_0^1 \frac{f\left(re^{2\pi\theta}\right)}{re^{2\pi\theta} - z_0} d\theta \tag{4}$$

this means that if we have uniformly random $\theta \in [0,1]$, then

$$I = f(z_0) = \mathbb{E}\left[\frac{f\left(re^{2\pi\theta}\right)}{re^{2\pi\theta} - z_0}\right] \tag{5}$$

the characteristic function of this random variable is

$$\mathbb{E}\left[e^{it\frac{f\left(re^{2\pi\theta}\right)}{re^{2\pi\theta}-z_0}}\right] \tag{6}$$

and we know that its derivative at zero has to equal I. Again by Cauchy integral theorem, the derivative of (6) wrt t evaluated at zero is

$$\int_0^1 e^{-2\pi i\xi}\mathbb{E}\left[e^{ie^{2\pi i\xi}\frac{f\left(re^{2\pi\theta}\right)}{re^{2\pi\theta}-z_0}}\right]d\xi=\int_0^1 \mathbb{E}\left[e^{i\left(\frac{e^{2\pi i\xi}f\left(re^{2\pi\theta}\right)}{re^{2\pi\theta}-z_0}-2\pi\xi\right)}\right]d\xi=I \tag{7}$$

which again means that the following expectation is our value in concern:

$$I=f\left(z_0\right)=\mathbb{E}_{\xi\in[0,1]}\mathbb{E}_{\theta\in[0,1]}\left[e^{i\left(\frac{e^{2\pi i\xi}f\left(re^{2\pi\theta}\right)}{re^{2\pi\theta}-z_0}-2\pi\xi\right)}\right] \tag{8}$$

We could obviously continue this way and more deeply iterated exponential.

We can generalize this approach to a wider family of compositions. Consider a function $k\left(x,y\right)$ holomorphic in y and denote by k_y its derivative wrt it second argument, also assume that

$$k_y\left(x,0\right)=x \tag{9}$$

then for all $r>0$

$$\int_{\mathcal{I}} f\left(t\right)dt=\int_{\mathcal{I}} k_y\left(f\left(t\right),0\right)dt=\frac{1}{2\pi i}\int_{\mathcal{I}}\int_0^1 r^{-1}e^{-2\pi\theta}k\left(f\left(t\right),re^{2\pi\theta}\right)d\theta dt \tag{10}$$

we observe that we can transform integrals in a compositional way, but unlike change-of-variable where the integrand is composed with some new function, on our case the integrand now becomes a variable inside an outer function.

Why would one want to do so? Not only we make the integrand much more complicated, but we also increase the dimension of integration as we began with a one-dimensional integral and ended up with a multivariable integral.

At the scope of numerical integration, we have mainly two methods: the Trapezoid rule (quite deterministic) and Monte-Carlo integration (randomized). Monte-Carlo integration's convergence, surprisingly, is independent on the dimension. So if this method is applied, we theoretically lose nothing from having multiple integrals.

For the trapezoid rule, it is known that its error can be calculated accurately using calculus of residues, that suggests good convergence for the trapezoid rule only when the function is well-bounded over the set $\{x+iy\}$ where $x\in\mathbb{R}$ and $|y|\leq r$ for some $r>1$, and the more r gets larger, the exponentially-better the convergence. Note that this holds for error of integrals over real intervals - still the behavior of the function on the complex plane completely determines the trapezoid rule's error. For this case, sometimes we'd like to precondition our integrand, and this might be done using (10).