

# Some notes on extending the SOAP kernel to multi-component systems

Jonathan Vandermause

(Dated: September 28, 2018)

## I. COMPUTING THE POWER SPECTRUM

We wish to efficiently compute the power spectrum

$$p_{n\nu l} = \sqrt{\frac{8\pi^2}{2l+1}} \sum_m c_{nlm} c_{\nu lm}^*, \quad (1)$$

where  $c_{nlm}$  are expansion coefficients determined by

$$\rho(\vec{r}) = \sum_{nlm} c_{nlm} g_{nl}(r) Y_{lm}(\hat{r}). \quad (2)$$

$\rho(\vec{r})$  is the atomic neighbor density

$$\rho(\vec{r}) = \sum_i \exp(-\alpha|\vec{r} - \vec{r}_i|^2) \quad (3)$$

and  $g_{nl}(r)$  are an orthonormal set of radial basis functions

$$\int dr r^2 g_{nl}(r) g_{n'l}(r) = \delta_{nn'}. \quad (4)$$

Note that we allow the radial basis set to depend on  $l$ .

From [1], we have the following exact formula for  $\rho(\vec{r})$ :

$$\rho(\vec{r}) = \sum_{ilm} c_{ilm}^i(r) Y_{lm}(\hat{r}) \quad (5)$$

where

$$c_{ilm}^i(r) = 4\pi \exp(-\alpha(r^2 + r_i^2)) i_l(2\alpha r r_i) Y_{lm}^*(\hat{r}_i). \quad (6)$$

Here  $i_l(x)$  is the modified spherical Bessel function of the first kind and is related to the modified Bessel function of the first kind  $I_l(x)$  according to

$$i_l(x) = \sqrt{\frac{\pi}{2x}} I_{l+1/2}(x). \quad (7)$$

Let  $\phi_{nl}(r)$  be a set of functions and let

$$g_{nl}(r) = \sum_{n'} U_{nn'} \phi_{n'}(r) \quad (8)$$

be an orthonormal basis constructed out of  $\phi_n(r)$ . Then

$$c_{nlm} = \sum_{in'} 4\pi Y_{lm}^*(\hat{r}_i) U_{nn'} I_{n'l}(r_i) \quad (9)$$

where

$$I_{n'l} \equiv \int_0^\infty dr r^2 \exp(-\alpha(r^2 + r_i^2)) i_l(2\alpha r r_i) \phi_{n'l}(r). \quad (10)$$

The integral identity [2]

$$\int_0^\infty t^{\nu+1} I_\nu(bt) \exp(-p^2 t^2) dt = \frac{b^\nu}{(2p^2)^{\nu+1}} \exp\left(\frac{b^2}{4p^2}\right) \quad (11)$$

suggests that we choose the functions

$$\phi_{nl}(r) = r^l \exp(-\beta_n r^2), \quad (12)$$

which gives

$$I_{nl}(r_i) = \frac{1}{4} \sqrt{\frac{\pi}{\alpha r_i}} \exp\left(-\frac{\alpha \beta_n r_i^2}{\alpha + \beta_n}\right) \frac{(\alpha r_i)^{l+1/2}}{(\alpha + \beta_n)^{l+3/2}}. \quad (13)$$

Defining

$$a_{nl}(r_i) \equiv \sum_{n'} U_{nn'} I_{n'l}(r_i), \quad (14)$$

we may write the expansion coefficients  $c_{nlm}$  as

$$c_{nlm} = \sum_i 4\pi a_{nl}(r_i) Y_{lm}^*(\hat{r}_i). \quad (15)$$

The power spectrum then takes the form

$$\begin{aligned} p_{n\nu l} &= \frac{32\pi^3 \sqrt{2}}{\sqrt{2l+1}} \sum_{mii'} a_{nl}(r_i) a_{\nu l}(r_{i'}) Y_{lm}^*(\hat{r}_i) Y_{lm}(\hat{r}_{i'}) \\ &= \frac{32\pi^3 \sqrt{2}}{\sqrt{2l+1}} \sum_{ii'} a_{nl}(r_i) a_{\nu l}(r_{i'}) \left( \sum_m Y_{lm}^*(\hat{r}_i) Y_{lm}(\hat{r}_{i'}) \right). \end{aligned} \quad (16)$$

By the addition theorem

$$\sum_m Y_{lm}(\hat{x}) Y_{lm}^*(\hat{y}) = \frac{2l+1}{4\pi} P_l(\hat{x} \cdot \hat{y}), \quad (17)$$

where  $P_l$  is the Legendre polynomial of degree  $l$ , the expression for the power spectrum becomes

$$\boxed{p_{n\nu l} = 8\pi^2 \sqrt{4l+2} \sum_{ii'} a_{nl}(r_i) a_{\nu l}(r_{i'}) P_l(\hat{r}_i \cdot \hat{r}_{i'})} \quad (18)$$

## II. COMPUTING THE DERIVATIVE OF THE POWER SPECTRUM

To predict forces, the derivative of the power spectrum with respect to the Cartesian coordinates of the central atom is needed. Let  $\xi_{ki}$  denote the  $k^{\text{th}}$  Cartesian component of atom  $i$  and  $\xi_{k0}$  denote the  $k^{\text{th}}$  Cartesian component of the central atom. The derivative of the power spectrum is given by

$$\boxed{\frac{\partial p_{nl}}{\partial \xi_{k0}} = 8\pi^2 \sqrt{4l+2} \sum_{ii'} \left( \frac{\partial a_{nl}(r_i)}{\partial r_i} \frac{\partial r_i}{\partial \xi_{k0}} a_{\nu l}(r_{i'}) P_l(\hat{r}_i \cdot \hat{r}_{i'}) + \right.} \quad (19)$$

$$a_{nl}(r_i) \frac{\partial a_{\nu l}(r_{i'})}{\partial r_{i'}} \frac{\partial r_{i'}}{\partial \xi_{k0}} P_l(\hat{r}_i \cdot \hat{r}_{i'}) +$$

$$\left. a_{nl}(r_i) a_{\nu l}(r_{i'}) \frac{\partial P_l(\hat{r}_i \cdot \hat{r}_{i'})}{\partial (\hat{r}_i \cdot \hat{r}_{i'})} \frac{\partial (\hat{r}_i \cdot \hat{r}_{i'})}{\partial \xi_{k0}} \right)$$

We now show how each of the derivatives appearing in this expression can be efficiently evaluated. First, note that

$$a'_{nl}(r_i) \equiv \frac{\partial a_{nl}}{\partial r_i} = \sum_{n'} U_{nn'} I'_{n'l}(r_i), \quad (20)$$

where

$$I'_{nl}(r_i) \equiv \frac{\partial I_{nl}}{\partial r_i} = \frac{\sqrt{\pi}\alpha}{4} \left( \frac{1}{\alpha r_i} \right)^{3/2} \frac{(\alpha r_i)^{l+1/2}}{(\alpha + \beta_n)^{l+5/2}} (-2r_i^2 \alpha \beta_n + l(\alpha + \beta_n)). \quad (21)$$

Next, note that

$$\frac{\partial r_i}{\partial \xi_{k0}} = -\frac{\xi_{ki}}{r_i}, \quad (22)$$

assuming the central atom is fixed at the origin. The derivative of the Legendre polynomial may be efficiently evaluated using the identity

$$\frac{dP_n(x)}{dx} = \frac{n}{x^2 - 1} (xP_n(x) - P_{n-1}(x)). \quad (23)$$

Finally, it is easy to show that

$$\frac{\partial (\hat{r}_i \cdot \hat{r}_j)}{\partial \xi_0} = \frac{-(\xi_{ki} + \xi_{kj})r_i r_j + (\hat{r}_i \cdot \hat{r}_j)(r_j^2 \xi_{ki} + r_i^2 \xi_{kj})}{r_i^2 r_j^2}, \quad (24)$$

again assuming that the central atom is fixed at the origin.

## III. COMPUTING THE DERIVATIVE OF THE SOAP KERNEL

The full SOAP kernel is the dot product of normalized power spectra raised to a positive power. In our approach, the covariance between local energies  $\epsilon_i$  is modelled by the SOAP

kernel,

$$\langle \epsilon_i \epsilon_j \rangle = K(\hat{p}_i, \hat{p}_j) = (\hat{p}_i \cdot \hat{p}_j)^\xi, \quad (25)$$

where  $\hat{p}_i$  denotes the normalized power spectrum of environment  $i$ . To model forces, the covariance of the partial derivatives of the local energies is needed:

$$\begin{aligned} \left\langle \frac{\partial \epsilon_i}{\partial \xi_{ik}}, \frac{\partial \epsilon_j}{\partial \xi_{jl}} \right\rangle &= \frac{\partial^2 \langle \epsilon_i \epsilon_j \rangle}{\partial \xi_{ik} \partial \xi_{jl}} \\ &= \frac{\partial^2 K(\hat{p}_i, \hat{p}_j)}{\partial \xi_{ik} \partial \xi_{jl}} \\ &= \sum_{\alpha\beta} \frac{\partial p_{i\alpha}}{\partial \xi_{ik}} \frac{\partial^2 K(\hat{p}_i, \hat{p}_j)}{\partial p_{i\alpha} \partial p_{j\beta}} \frac{\partial p_{j\beta}}{\partial \xi_{jl}}, \end{aligned} \quad (26)$$

where  $p_{i\alpha}$  is component  $\alpha$  of the normalized power spectrum  $\hat{p}_i$ . The derivatives of the power spectra with respect to Cartesian components may be evaluated using the expressions from Section 2. The remaining term is

$$\begin{aligned} \frac{\partial^2 K(\hat{p}_i, \hat{p}_j)}{\partial p_{i\alpha} \partial p_{j\beta}} &= \frac{\partial}{\partial p_{i\alpha}} (\xi (\hat{p}_i \cdot \hat{p}_j)^{\xi-1} p_{i\beta}) \\ &= \xi(\xi-1) (\hat{p}_i \cdot \hat{p}_j)^{\xi-2} p_{j\alpha} p_{i\beta} + \xi (\hat{p}_i \cdot \hat{p}_j)^{\xi-1} \delta_{\alpha\beta} \\ &= \xi (\hat{p}_i \cdot \hat{p}_j)^{\xi-2} ((\xi-1) p_{j\alpha} p_{i\beta} + (\hat{p}_i \cdot \hat{p}_j) \delta_{\alpha\beta}). \end{aligned} \quad (27)$$

Note that if  $\xi = 1$ , the double sum collapses into a single sum. Raising the SOAP kernel to a power  $\xi > 1$  therefore comes at a significant computational cost. Moreover, the distance between environments can be "sharpened"

#### IV. REVIEW OF THE SINGLE-COMPONENT SOAP KERNEL

In the SOAP scheme, an atomic environment is represented by an atomic neighbor density function  $\rho(\vec{r})$ , where

$$\rho(\vec{r}) = \sum_i \exp(-\alpha |\vec{r} - \vec{r}_i|^2), \quad (28)$$

$\alpha$  is a hyperparameter, and  $\vec{r}_i$  is the position of atom  $i$  in the environment of the central atom. The single-component SOAP kernel between two atomic neighbor densities  $\rho$  and  $\rho'$  is defined by Eq. (36) of [1] as:

$$K(\rho, \rho') = \left( \frac{k(\rho, \rho')}{\sqrt{k(\rho, \rho) k(\rho', \rho')}} \right)^\xi, \quad (29)$$

where  $k(\rho, \rho')$  is a rotationally invariant symmetry kernel given by

$$k(\rho, \rho') = \int d\hat{R} \left| \int \rho(\vec{r}) \rho'(\hat{R}\vec{r}) d\vec{r} \right|^n, \quad (30)$$

and  $\hat{R}$  is a 3-D rotation. The outer integral is over all rotations and ensures that the kernel is rotationally symmetric.

In Section IV(B) of [1], it is shown that  $k(\rho, \rho')$  is equal to the dot product of the "rotational power spectra" of densities  $\rho$  and  $\rho'$ , which is efficient to compute. I will review this derivation here, since a similar argument can be applied to multi-component systems.

Letting  $g_n(r)$  be an orthonormal set of radial basis functions, so that

$$\int dr r^2 g_n(r) g_{n'}(r) = \delta_{nn'}, \quad (31)$$

an atomic neighbor density  $\rho(\vec{r})$  may be expanded into spherical harmonics as

$$\rho(\vec{r}) = \sum_{nlm} c_{nlm} g_n(r) Y_{lm}(\vec{r}). \quad (32)$$

We first compute that

$$\begin{aligned} S(\rho, \hat{R}\rho') &\equiv \int d\vec{r} \rho(\vec{r}) \rho'(\hat{R}\vec{r}) \\ &= \int d\vec{r} \left( \sum_{nlm} c_{nlm}^* g_n(r) Y_{lm}^*(\hat{r}) \right) \left( \sum_{n'l'm'} c'_{n'l'm'} g_{n'}(r) Y_{l'm'}(\hat{R}\vec{r}) \right) \\ &= \sum_{nlmn'l'm'} c_{nlm}^* c'_{n'l'm'} \int dr r^2 g_n(r) g_{n'}(r) \int d\hat{r} Y_{lm}^*(\hat{r}) Y_{l'm'}(\hat{R}\vec{r}) \\ &= \sum_{nlml'm'} c_{nlm}^* c'_{nl'm'} \int d\hat{r} Y_{lm}^*(\hat{r}) Y_{l'm'}(\hat{R}\vec{r}). \end{aligned} \quad (33)$$

Note that

$$\begin{aligned} Y_{l'm'}(\vec{R}\hat{r}) &= \langle \vec{r} | \hat{R}^\dagger | Y_{l'm'} \rangle \\ &= \sum_{m''} \langle \vec{r} | Y_{l'm''} \rangle \langle Y_{l'm''} | \hat{R}^\dagger | Y_{l'm'} \rangle \\ &= \sum_{m''} \langle \vec{r} | Y_{l'm''} \rangle \langle Y_{l'm'} | \hat{R} | Y_{l'm''} \rangle^* \\ &= \sum_{m''} Y_{l'm''}(\hat{r}) D^*(\hat{R})_{m'l'm'}, \end{aligned} \quad (34)$$

where the Wigner function  $D(\hat{R})_{mm'}^l$  is defined as

$$D(\hat{R})_{mm'}^l = \langle Y_{lm} | \hat{R} | Y_{lm'} \rangle. \quad (35)$$

Plugging into Eq. (33) gives

$$\begin{aligned}
S(\rho, \hat{R}\rho') &= \sum_{nlml'm'm''} c_{nlm}^* c'_{nl'm'} D^*(\hat{R})_{m'm''}^{l'} \int d\hat{r} Y_{lm}^*(\hat{r}) Y_{l'm''}(\hat{r}) \\
&= \sum_{nlmm'} c_{nlm}^* c'_{nlm'} D^*(\hat{R})_{m'm}^l.
\end{aligned} \tag{36}$$

The symmetry kernel  $k(\rho, \rho')$  with  $n = 2$  then takes the form

$$\begin{aligned}
k(\rho, \rho') &= \int d\hat{R} S(\rho, \hat{R}\rho') S^*(\rho, \hat{R}\rho') \\
&= \sum_{nlmm'n'\lambda\mu\mu'} c_{nlm}^* c'_{nlm'} c_{n'\lambda\mu} (c'_{n'\lambda\mu'})^* \int d\hat{R} D^*(\hat{R})_{m'm}^l D(\hat{R})_{\mu'\mu}^{\lambda} \\
&= \sum_{nlmm'n'} \frac{8\pi^2}{2l+1} c_{nlm}^* c'_{nlm'} c_{n'lm} (c'_{n'lm'})^* \\
&= \sum_{nn'l} \frac{8\pi^2}{2l+1} \left( \sum_m c_{nlm}^* c_{n'lm} \right) \left( \sum_{m'} c'_{nlm'} (c'_{n'lm'})^* \right) \\
&= \sum_{nn'l} \frac{8\pi^2}{2l+1} p_{n'nl} p'_{nn'l},
\end{aligned} \tag{37}$$

where  $p_{nn'l} \equiv \sum_m c_{nlm} c_{n'lm}^*$  is the rotational power spectrum. The power spectrum of a real function is real, so  $p_{n'nl} = p_{nn'l}$ , and  $k(\rho, \rho')$  may be expressed as the dot product of the power spectra  $\vec{p}$  and  $\vec{p}'$ :

$$k(\rho, \rho') = \vec{p} \cdot \vec{p}'. \tag{38}$$

The full SOAP kernel  $K(\rho, \rho')$  is the dot product of normalized power spectra raised to a positive power:

$$K(\rho, \rho') = \left( \frac{\vec{p}}{|\vec{p}|} \cdot \frac{\vec{p}'}{|\vec{p}'|} \right)^\xi. \tag{39}$$

## V. REVIEW OF THE MULTI-COMPONENT SOAP KERNEL

We now review the multi-species version of the SOAP kernel proposed in [3]. Consider a two-component system. Let  $\rho$  and  $\rho'$  denote two atomic environments, with  $\rho_A, \rho_B$  the atomic neighbor densities of atom types  $A$  and  $B$  in environment  $\rho$ , and  $\rho'_A, \rho'_B$  the densities of  $A$  and  $B$  in environment  $\rho'$ . We define the MOP kernel between  $\rho$  and  $\rho'$  by direct analogy to the SOAP kernel:

$$K_{\text{MOP}}(\rho, \rho') = \left( \frac{k_{\text{MOP}}(\rho, \rho')}{\sqrt{k_{\text{MOP}}(\rho, \rho) k_{\text{MOP}}(\rho', \rho')}} \right)^\xi, \tag{40}$$

where  $k_{\text{MOP}}(\rho, \rho')$  is a rotationally invariant symmetry kernel given by

$$\begin{aligned}
k_{\text{MOP}}(\rho, \rho') &= \int d\hat{R} \left| \int \left( \rho_A(\vec{r}) \rho'_A(\hat{R}\vec{r}) + \rho_B(\vec{r}) \rho'_B(\hat{R}\vec{r}) \right) d\vec{r} \right|^2 \\
&= \int d\hat{R} \left| \int \rho_A(\vec{r}) \rho'_A(\hat{R}\vec{r}) d\vec{r} \right|^2 + \int d\hat{R} \left| \int \rho_B(\vec{r}) \rho'_B(\hat{R}\vec{r}) d\vec{r} \right|^2 \\
&\quad + \int d\hat{R} \left( \int d\vec{r} \rho_A(\vec{r}) \rho'_A(\hat{R}\vec{r}) \right) \left( \int d\vec{r} \rho_B(\vec{r}) \rho'_B(\hat{R}\vec{r}) \right)^* \\
&\quad + \int d\hat{R} \left( \int d\vec{r} \rho_B(\vec{r}) \rho'_B(\hat{R}\vec{r}) \right)^* \left( \int d\vec{r} \rho_A(\vec{r}) \rho'_A(\hat{R}\vec{r}) \right) \\
&= k(\rho_A, \rho'_A) + k(\rho_B, \rho'_B) + k_{\text{cross}}(\rho, \rho') + k_{\text{cross}}^*(\rho, \rho'),
\end{aligned} \tag{41}$$

which is the sum of the SOAP kernels between the two density types plus a cross term  $k_{\text{cross}}$ , which we now analyze in detail.

From the derivation in part 1, we have that

$$\int d\vec{r} \rho_A(\vec{r}) \rho'_A(\hat{R}\vec{r}) = \sum_{nlmm'} (c_{nlm}^A)^* c_{nlm}'^A D^*(\hat{R})_{m'm}^l, \tag{42}$$

$$\int d\vec{r} \rho_B(\vec{r}) \rho'_B(\hat{R}\vec{r}) = \sum_{n'\lambda\mu\mu'} (c_{n'\lambda\mu}^B)^* c_{n'\lambda\mu}'^B D^*(\hat{R})_{\mu'\mu}^l, \tag{43}$$

and thus

$$\begin{aligned}
k_{\text{cross}}(\rho, \rho') &= \int d\hat{R} S(\rho_A, \hat{R}\rho'_A) S^*(\rho_B, \hat{R}\rho'_B) \\
&= \sum_{nlmm'n'\lambda\mu\mu'} (c_{nlm}^A)^* c_{nlm}'^A c_{n'\lambda\mu}^B (c_{n'\lambda\mu'}^B)^* \int d\hat{R} D^*(\hat{R})_{m'm}^l D(\hat{R})_{\mu'\mu}^\lambda \\
&= \sum_{nlmm'n'} \frac{8\pi^2}{2l+1} (c_{nlm}^A)^* c_{nlm}'^A c_{n'lm}^B (c_{n'lm'}^B)^* \\
&= \sum_{nn'l} \frac{8\pi^2}{2l+1} \left( \sum_m (c_{nlm}^A)^* c_{n'lm}^B \right) \left( \sum_m c_{nlm}'^A (c_{n'lm'}^B)^* \right) \\
&= \vec{p}_{AB} \cdot \vec{p}_{AB}',
\end{aligned} \tag{44}$$

which is the dot product of the cross power spectra of the two environments. The symmetric MOP kernel  $k_{\text{MOP}}(\rho, \rho')$  before normalization therefore takes the form

$$k_{\text{MOP}}(\rho, \rho') = \vec{p}_A \cdot \vec{p}_A' + \vec{p}_B \cdot \vec{p}_B' + 2\vec{p}_{AB} \cdot \vec{p}_{AB}'. \tag{45}$$

Notice that if one of the densities vanishes, e.g.  $\vec{p}_A = 0$ , the MOP kernel reduces to the SOAP kernel. MOP may therefore be viewed as a generalization of SOAP that is suitable



for multicomponent systems.

---

- [1] A. P. Bartók, R. Kondor, and G. Csányi, *Physical Review B* **87**, 184115 (2013).
- [2] Digital Library of Mathematical Functions: <https://dlmf.nist.gov/10.43>.
- [3] S. De, A. P. Bartók, G. Csányi, and M. Ceriotti, *Physical Chemistry Chemical Physics* **18**, 13754 (2016).