Holographic Learning of Protein Structures

Michael Pun

21 May 2020

Contents

1 Protein background

1.1 Functions

1.2 Structure

1.2.1 Secondary Structure

| Table 1.2	Torsion angles for regular polypeptide conformations	
	Torsion angle (degrees)	
Structure	ϕ	ψ
Hypothetical fully extended	$+180\ (\equiv -180)$	$+180$
Antiparallel β sheet	-139	$+135$
Parallel β sheet	-119	$+113$
Right-handed α helix	-57	-47
Left-handed α helix	$\sim +60$	$\sim +60$
3_{10} helix	-49	-26
π helix	-57	-70

Note that $+180°$ is the same as $-180°$.
From G. N. Ramachandran and V. Sasisekharan, *Adv. Protein Chem.* **23**, 283 (1968). IUPAC-IUB
Commission on Biochemical Nomenclature, *Biochemistry* **9**, 3471 (1970).

2   Holography

2.1   In Line Holography
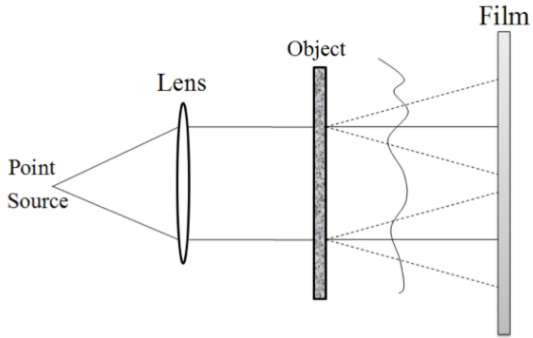
holography.png

Typical holography involves shiniing a coherent monochromatic light source on an object and recording in intensity of an interference pattern that the re-emitted light makes with a reference beam. In Figure 1, the object is semi-transparent so the reference wave is the source wave that is not absorbed by the object.?

The holographic film is assumed to have a transmission function that is proportional to the intensity of the light recieved $t(x, y) = a + bI(x, y)$ where $I(x, y) = |U_r + U_o(x, y)|^2$.

This means when the film is illuminated with another coherent monochromatic beam U_0, the light transmitted is given by

$$U(x, y) = U_0 t(x, y) \tag{1}$$
$$= U_0(a + b|U_0|^2) + bU_0|U_o(x, y)|^2 + b|U_0|^2 U_o(x, y) + bU_0^2 U_o^*(x, y) \tag{2}$$

The last two terms contain the field information due to the object and by uniqueness of Maxwell's equations, they imply two images of the object–one virtual and one real.

The takeaway for our purposes is that the reference beam is used simply to create an interference pattern that can then recover the full complex information of the original field.

If we can simply write down the complex field on the recording surface then we are effectively doing holography.

2.2   Spherical holography

The idea of spherical holography is to use the holographic framework to project the information contained in 3-dimensional point cloud onto a 2-dimensional surface in a rotationally covariant way.

The use of planar waves and a planar recording surface breaks spherical symmetry. In spherical symmetry, we use a spherical reference wave and a spherical recording surface.

We can avoid choosing a preferred direction; however we must choose an origin for our spherical waves and recording surface. Once this point is chosen, a spherical wave is emmitted from the origin and travels outward exciting the points in our point cloud individually. These excited points then re-emit a spherical wave with the same frequency.

We assume enough time has passed so all points have been radiating for long enough so that the wavefronts from all points have completely passed the spherical recording surface. The complete complex electric field is recorded at this time ignoring the reference wave. This recording is our spherical hologram.
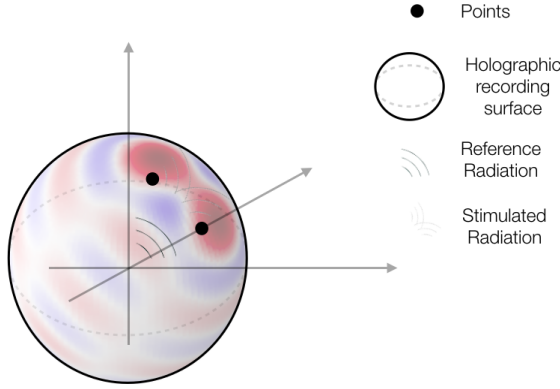

Figure: Illustration to show the mechanics of spherical holography.

Spherical wave expansion

$$\frac{e^{ik|\mathbf{r}-\mathbf{r}'|}}{|\mathbf{r}-\mathbf{r}'|} = ik \sum_l j_l(kr_<)h_l^{(1)}(kr_>) \sum_m Y_{lm}^*(\theta',\phi')Y_{lm}(\theta,\phi). \tag{3}$$

3 Group Theory

3.1 Wigner D-matrices are irreps of SO(3)

We know from study of SO(2) that the rotations about an axis $\hat{n}$ are generated by J_n. In other words,

$$R_n(\theta) = e^{-i\theta J_n}.$$

Meanwhile we can parametrize rotations in SO(3) by the Euler angles and express elements of SO(3) as

$$R(\alpha,\beta,\gamma) = R_3(\alpha)R_2(\beta)R_3(\gamma) \tag{4}$$

$$= e^{-i\alpha J_3}e^{-i\beta J_2}e^{-i\gamma J_3} \tag{5}$$

Now if we can find a basis which is simultaneous eigenstates we would already have found our representation. Unfortunately J_2 and J_3 do not commute. Thus we need to look for a different basis.

The operator J^2 commutes with all generators of SO(3). We will eventually use a simultaneous basis of J^2 and J_3 as our basis for the irreducible representations.

Since J^2 commutes with all generators of our group algebra, Schur's first lemma tells us that J^2 is a multiple of the identity in any irreducible representation.

Meanwhile we note that the commutator of any two generators is

$$[J_i, J_j] = i\epsilon_{ijk}J_k \tag{6}$$

This statement is equivalent to the group multiplication rule.

We also define

$$J_{\pm} = J_1 \pm iJ_2$$

which will be a raising (lowering) operator. Using commutation relations and identities we see that eigenvectors of J_3 must have eigenvalues given by $m = -j, -j+1, ..., j-1, j$ where j is related to eigenvalue of J^2. Knowing that the raising and lowering operators raise and lower the eigenvalue of J_3 by 1 at each application, we get that $j = 0, 1/2, 1, 3/2, ...$ This allows us to use the basis $|jm\rangle$ to see our irreducible representations are labeled by j.

Specifically these basis vectors transform under rotations as

$$U(\alpha, \beta, \gamma) |jm\rangle = |jm'\rangle D^j(\alpha, \beta, \gamma)^{m'}_m \tag{7}$$

where the Wigner D-matrices are given by

$$D^j(\alpha, \beta, \gamma)^{m'}_m = e^{-i\alpha m'} d^j(\beta)^{m'}_m e^{-i\gamma m'} \tag{8}$$

where

$$d^j(\beta)^{m'}_m = \langle jm' | e^{-i\beta J_2} |jm\rangle \tag{9}$$

3.2 Wigner D-matrices form complete orthogonal basis over functions from SO(3) $\to \mathbb{C}$

Orthonormality:

$$(2j+1) \int d\tau_R D^{\dagger}_j(R)^m_n D^{j'}(R)^{n'}_{m'} = \delta^{j'}_j \delta^{n'}_n \delta^m_{m'} \tag{10}$$

Completeness:

$$\sum_{jmn} (2j+1)D^{\dagger}_j(R')^m_n D^j(R)^n_m = \delta(R - R') \tag{11}$$

3.3 Wigner D-matrices naturally decompose $\mathbb{C}(SO(3))$ into invariant subspaces via the projection operators

Suppose we have a state of fixed "linear momentum" $|p\rangle$ and with spin λ

3.3.1 Mike's convolution

Given two functions $f, h : SO(3) \to \mathbb{C}$, the convolution is

$$(h \star f)(R) = \int dS f^*(S) h(S^{-1}R) \tag{12}$$

The fourier coefficient of a convolution is

$$(h \star f)^l_{mm'} = \int dR D^\dagger_l(R)_{mm'} \int dS f^*(S) h(S^{-1}R) \tag{13}$$

$$= \int dS D^\dagger_l(S)_{mn} f^*(S) \int dR D^\dagger_l(S^{-1}R)_{nm'} g(S^{-1}R) \tag{14}$$

$$= f^l_{mn} g^l_{nm'} \tag{15}$$

3.3.2 Convolution for functions over S^2

Let $f, h : S^2 \to \mathbb{C}$. The convolution is then

$$(h \star f)(R) = \int d\Omega f(\hat{n}) h(R^{-1}\hat{n}) \tag{16}$$

$$(h \star f)^l_{mn} = \int dR D^\dagger_l(R)_{mn} \int d\Omega \sum_{l'm'} f^*_{l'm'} Y^*_{l'm'}(\hat{n}) \sum_{l''m''} h_{l''m''} D_{l''}(R)_{n''m''} Y_{l''n''}(\hat{n}) \tag{17}$$

$$= \int d\Omega \sum_{l'm'} \sum_{l''m''} f^*_{l'm'} h_{l''m''} Y^*_{l'm'}(\hat{n}) Y_{l''n''}(\hat{n}) \underbrace{\int dR D^\dagger_l(R)_{mn} D_{l''}(R)_{n''m''}}_{(2l+1)^{-1}\delta_{ll''}\delta_{mn''}\delta_{nm''}} \tag{18}$$

$$= (2l+1)^{-1} \sum_{l'm'} f^*_{l'm'} h_{ln} \underbrace{\int d\Omega Y^*_{l'm'}(\hat{n}) Y_{lm}(\hat{n})}_{\delta_{l'l}\delta_{m'm}} \tag{19}$$

$$= (2l+1)^{-1} f^*_{lm} h_{ln} \tag{20}$$

$$\tag{21}$$

4 Covariant network

The covariant network is the Clebsch-Gordan Network presented in **?**. The network takes in information classified by representation of SO(3). It then performs two operations within one layer of the network. First, a linear combination is performed where only information of corresponding to the same representation is combined. Second, a non-linearity is taken between all representation coefficients. While this non-linearity is generally a linear combination of multiple representations, it can be decomposed into a sum of representations using the Clebsch-Gordan coefficients.

The following are the details of how my network is currently running.

4.1 Inputs

The input to the network is a collection of fourier coefficients $f^0_{lc^l_0 m}$ where l indexes the representation and c indexes the channel. The index m is a function of l and can take on $2l + 1$ values.

4.2 Linear Combination

The first step to transform layer zero into layer one is to take a linear combination of the inputs. In general, this linear combination can either be a dimensional reduction or expansion.

Let $W^i_{c_i c^l_{i-1}}$ be a set of learnable weights. The parameter c_i is the dimension we want to expand or reduce the input into. The linear combination is then

$$g^i_{lc_i m} = W^i_{c_i c^l_{i-1}} f^{i-1}_{lc^l_{i-1} m} \tag{22}$$

4.3 Non-linearity Clebsch-Gordan Decomposition

The next step from layer to layer is taking a non-linearity of the linear combination. This non-linearity can be decomposed with Clebsch-Gordan coefficients which can be grouped into two dimensional matrices $C^l_{l_1 l_2}$ of size $(2l_1 + 1) \times (2l_2 + 1)$.

The non-linear product is then

$$f^i_{lc_i m} = (C^l_{l_1 l_2})_{m_1 m_2} g^i_{l_1 c_i m_1} g^i_{l_2 c_i m_2} \tag{23}$$

4.4 Fully Connected Layer

For classification tasks, the scalar information of the last layer is used as the input into a fully connected layer. For an n layer network with p classes, this operation looks like

$$f^{n+1}_k = W^{n+1}_{kc^0_{n-1}} f^n_{0c^0_n} \tag{24}$$

where k takes on p possible values. f^{n+1}_k are the log probabilities of each of the p classes.

5   Results

5.1   Classification of Amino Acid Structures
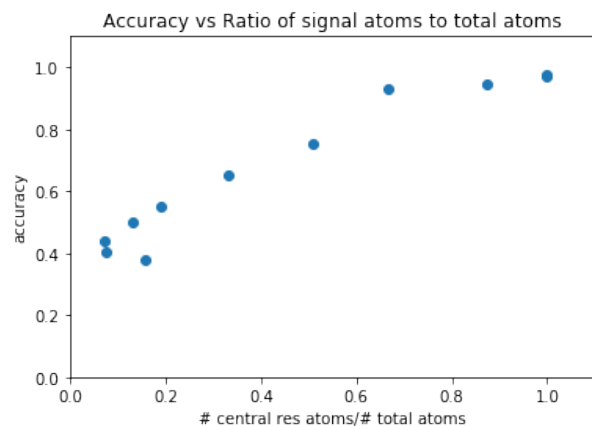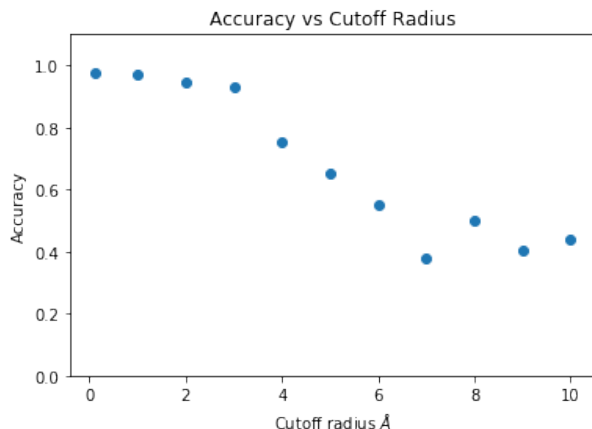
5.2   Classification of neighborhoods

5.2.1   Classification of neighborhoods restricted to secondary structure

References

Gabriel   Popescu.   *Holography*.   http://light.ece.illinois.edu/ECE460/PDF/Hologra-phy.pdf

Risi Kondor, Zhen Lin, Shubhendu Trivedi. Clebsch-Gordan Nets: A Fully Fourier Space Spherical Convolution Neural Network arXiv:1806.09231v2