Supplementary Material for
“On the Number of Linear Functions Composing Deep Neural Networks: Towards a Refined Definition of Neural Network Complexity”

Yuuki Takai
RIKEN AIP
yuuki.takai@riken.jp

Akiyoshi Sannai
RIKEN AIP
akiyoshi.sannai@riken.jp

Matthieu Cordonnier
École Normale Supérieure Paris-Saclay
matthieu.cordonnier@ens-paris-saclay.fr

A Illustrations and examples

A.1 Fully connected shallow model

In this section, we illustrate linear region calculations for simple examples in the plane.

In the two-dimensional plane, "general position" means that two lines always intersect and three lines never are concurrent. Let us see an example in the case $n_0 = 2$, $n_1 = 4$, i.e., four lines in the plan.

$$\begin{cases} x - y + 1 = 0 & : \ H_1 \\ x - y - 1 = 0 & : \ H_2 \\ x + y - 2 = 0 & : \ H_3 \\ x = \frac{1}{2} & : \ H_4. \end{cases}$$

This arrangement is not in general position because H_1 and H_2 are parallel or H_1, H_3 and H_4 are concurrent. Its number of chambers is 9 (Figure 1)

Let us modify H_2 to make the arrangement being general position. Now we have:

$$\begin{cases} x - y + 1 = 0 & : \ H_1 \\ y = 1 & : \ H_2 \\ x + y - 2 = 0 & : \ H_3 \\ x = \frac{1}{2} & : \ H_4. \end{cases}$$

Now the number of chambers is $11 = \sum_{i=0}^{n_0} \binom{n_1}{i}$. It is maximal for a 4-line arrangement in the real plane (Figure 2).

A.2 Permutation invariant shallow model

Let us consider an example of a permutation-invariant shallow model with $m = n = 2$, i.e., this model also implements a function from $\mathbb{R}^2$ to $\mathbb{R}^4$. We have the two pairs of lines (Figure 3):

$$\begin{cases} 2x + \frac{1}{2}y - 3 = 0 & : \ H_{11} \\ \frac{1}{2}x + 2y - 3 = 0 & : \ H_{12} \\ -x + 6y = 0 & : \ H_{21} \\ 6x - y = 0 & : \ H_{22}. \end{cases}$$

We also count 11 chambers.

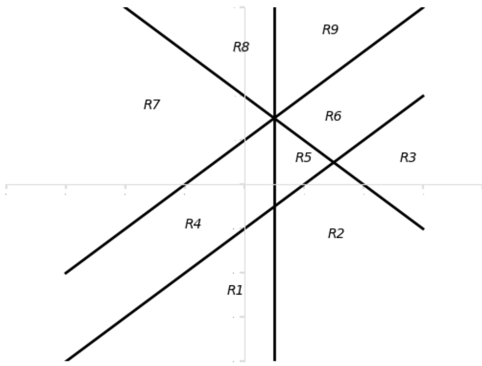

Figure 1: The line arrangement not in general position. The number of chambers is 9.

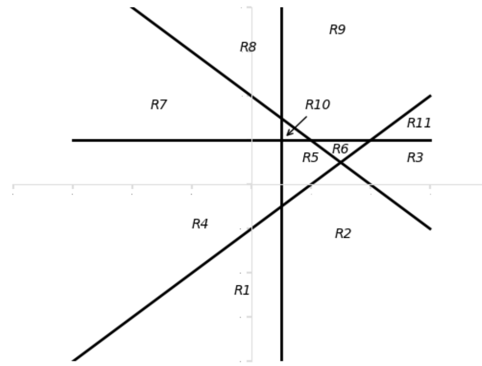

Figure 2: The line arrangement in general position. The number of chambers is 11 and is maximal

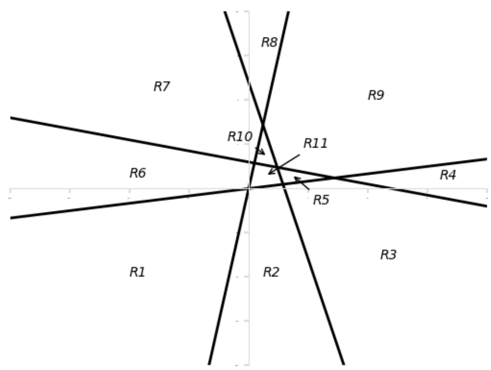

Figure 3: The 4 lines arrangement in the plane of a permutation invariant model. We count 11 linear regions.

A.3 Measure of complexity as the number of equivalent classes

Let us consider again the last invariant model example:

$$\begin{cases} 2x + \frac{1}{2}y - 3 = 0 & : H_{11} \\ \frac{1}{2}x + 2y - 3 = 0 & : H_{12} \\ -x + 6y = 0 & : H_{21} \\ 6x - y = 0 & : H_{22}. \end{cases}$$

In this case, S_2 has a single element which is the permutation $\sigma = (1\ 2)$. Here, the action of σ on $\mathbb{R}^2$ is exactly the action of the reflection symmetry through the line $x = y$. Then, the corresponding Euclidean transformation ϕ is $\phi = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$ and the underlying group is $\hat{\Phi} = \{I, \phi\}$

In Figure 4, we identify regions belonging to the same equivalent classes. In this case, a region is identified by its symmetry through the line $x = y$. Therefore, we count 7 equivalent classes of linear regions: {R1}, {R2,R6}, {R3,R7}, {R4,R8}, {R5,R10}, {R9}, {R11}.

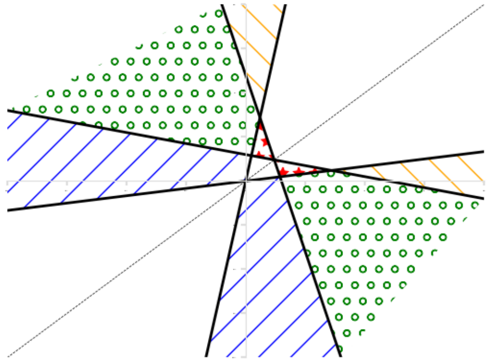

Figure 4: The dashed line is the line of equation $x = y$. We identify the equivalent regions with respect to the symmetry through the line $x = y$. The number of orbits is 7.

B Proof of Proposition 1

In this section, we prove Proposition 1. To show this, we use the Deletion-Restriction theorem (Orlik and Terao, 2013, Theorem 2.56 and Theorem 2.68).

Theorem 1 (Brylawsky, Zaslavsky)**.** *For a hyperplane arrangement $\mathcal{A}$ in $\mathbb{R}^n$ and a fixed hyperplane $X \in \mathcal{A}$, let $(\mathcal{A}, \mathcal{A}', \mathcal{A}'')$ be the triple defined as $\mathcal{A}' = \mathcal{A} \backslash \{X\}$ and*

$$\mathcal{A}'' = \{H \cap X \mid H \in \mathcal{A} \backslash \{X\}, H \cap X \neq \emptyset\}.$$

Then, the following holds:

$$|\mathrm{Ch}(\mathcal{A})| = |\mathrm{Ch}(\mathcal{A}')| + |\mathrm{Ch}(\mathcal{A}'')|.$$

By apply Theorem 1 to our hyperplane arrangement, we obtain a recurrence relation and calculate the number of linear regions for permutation invariant models.

Proof of Proposition 1. Let $\mathcal{B}_{m,n} = \{H_{ij} \subset \mathbb{R}^n \mid i = 1, \ldots, m, \ j = 1, \ldots, n\}$ be the hyperplane arrangement defined by (2.6). We recall that hyperplanes of this arrangement $\mathcal{B}_{m,n}$ satisfy the following equations:

$$H_{i_1,j} \cap H_{i_2,j} \cap H_{i_3,j} = \emptyset, \tag{B.1}$$

$$H_{i_1,j_1} \cap H_{i_1,j_2} \cap H_{i_2,j_1} = H_{i_1,j_1} \cap H_{i_1,j_2} \cap H_{i_2,j_2} = H_{i_1,j_1} \cap H_{i_2,j_1} \cap H_{i_2,j_2} \tag{B.2}$$

for $i_1, i_2, i_3 = 1, \ldots, m$ and $j, j_1, j_2 = 1, \ldots, n$.

We apply Theorem 1 to $\mathcal{B}_{m,n}$ and $H_{m,n} \in \mathcal{B}_{m,n}$. Then, we have

$$\mathcal{B}'_{m,n} = \{H_{11}, \ldots, H_{1n}, \ldots, H_{m1}, \ldots, H_{m,n-1}\},$$
$$\mathcal{B}''_{m,n} = \{H_{11} \cap H_{m,n}, \ldots, H_{1n} \cap H_{m,n}, \ldots, H_{m1} \cap H_{m,n}, \ldots, H_{m,n-1} \cap H_{m,n}\}$$

and $|\mathcal{B}_{m,n}| = |\mathcal{B}'_{m,n}| + |\mathcal{B}''_{m,n}|$ Here, because $H_{m,n}$ is a hyperplane bijective to $\mathbb{R}^{n-1}$, $H_{ij} \cap H_{m,n}$ can be regarded as a hyperplane in $H_{m,n} = \mathbb{R}^{n-1}$.

Next, we consider deletion and restriction for $\mathcal{B}''_{m.n}$ and $H_{m-1,n} \cap H_{m,n}$. Then, we have

$$(\mathcal{B}''_{m,n})' = \left\{ \begin{array}{c} H_{1,1} \cap H_{m,n}, \ldots, H_{m-2,n} \cap H_{m,n}, H_{m-1,1} \cap H_{m,n}, \ldots, H_{m-1,n-1} \cap H_{m,n}, \\ H_{m,1} \cap H_{m,n}, \ldots, H_{m,n-1} \cap H_{m,n} \end{array} \right\},$$

$$(\mathcal{B}''_{m,n})'' = \left\{ \begin{array}{c} H_{1,1} \cap H_{m-1,n} \cap H_{m,n}, \ldots, H_{1,n} \cap H_{m-1,n} \cap H_{m,n}, \ldots, \\ H_{m,1} \cap H_{m-1,n} \cap H_{m,n}, \ldots, H_{m,n-1} \cap H_{m-1,n} \cap H_{m,n} \end{array} \right\}.$$

Then, in the above $(\mathcal{B}''_{m,n})''$, by the relation (B.3), we have

$$H_{i,n} \cap H_{m-1,n} \cap H_{m,n} = \emptyset$$

for any $i = 1, \ldots, m-2$. Hence, any hyperplane of the form $H_{i,n} \cap H_{m-1,n} \cap H_{m,n}$ vanishes from $(\mathcal{B}''_{m,n})''$. Moreover, by the relation (B.4), for any $j = 1, \ldots, n-1$,

$$H_{m,j} \cap H_{m-1,n} \cap H_{m,n} = H_{m-1,j} \cap H_{m-1,n} \cap H_{m,n}$$

holds. By this relation, we can unify the hyperplanes of forms of $H_{m,j} \cap H_{m-1,n} \cap H_{m,n}$ and $H_{m-1,j} \cap H_{m-1,n} \cap H_{m,n}$. By these arguments, $(\mathcal{B}''_{m,n})''$ can be written by

$$(\mathcal{B}''_{m,n})'' = \{H_{i,j} \cap H_{m-1,n} \cap H_{m,n} \subset \mathbb{R}^{n-2} \mid i = 1, \ldots, m-1,\ j = 1, \ldots, n-1\}.$$

Once, we set $\overline{H}_{i,j} = H_{i,j} \cap H_{m-1,n} \cap H_{m,n} \in (\mathcal{B}''_{m,n})''$. Then, it is easy to show that the obtained arrangement $(\mathcal{B}'')'' = \{\overline{H}_{i,j} \subset \mathbb{R}^{n-2} \mid i = 1, \ldots, m-1, j = 1, \ldots, n-1\}$ satisfies the following relations:

$$\begin{aligned}
&\overline{H}_{i_1,j} \cap \overline{H}_{i_2,j} \cap \overline{H}_{i_3,j} = \emptyset,\\
&\overline{H}_{i_1,j_1} \cap \overline{H}_{i_1,j_2} \cap \overline{H}_{i_2,j_1} = \overline{H}_{i_1,j_1} \cap \overline{H}_{i_1,j_2} \cap \overline{H}_{i_2,j_2} = \overline{H}_{i_1,j_1} \cap \overline{H}_{i_2,j_1} \cap \overline{H}_{i_2,j_2}.
\end{aligned}$$

for $i_1, i_2, i_3 = 1, \ldots, m-1$ and $j, j_1, j_2 = 1, \ldots, n-1$. This means that the hyperplane arrangement $(\mathcal{B}'')''$ can be regarded as an arrangement "$\mathcal{B}_{m-1,n-1}$ in $\mathbb{R}^{n-2}$". We will subsequently justify this argument more precisely.

Before we do it, we shall observe the deletion and restriction for $\mathcal{B}'_{m,n}$ with $H_{m-1,n} \in \mathcal{B}'_{m,n}$. Then, we have the following arrangements:

$$\begin{aligned}
(\mathcal{B}'_{m,n})' &= \{H_{1,1}, \ldots, H_{m-2,n}, H_{m-1,1}, \ldots, H_{m-1,n-1}, \ldots, H_{m,1}, \ldots, H_{m,n-1}\},\\
(\mathcal{B}'_{m,n})'' &= \left\{ \begin{array}{c} H_{11} \cap H_{m-1,n}, \ldots, H_{1,n} \cap H_{m-1,n}, \ldots, H_{m-1,1} \cap H_{m-1,n}, \ldots, \\ H_{m-1,n-1} \cap H_{m-1,n}, H_{m,1} \cap H_{m-1,n}, \ldots, H_{m,n} \cap H_{m-1,n} \end{array} \right\}.
\end{aligned}$$

Then, we remark that $(\mathcal{B}'_{m,n})''$ is same as $(\mathcal{B}''_{m,n})'$ if we exchange $H_{m-1,j}$ and $H_{m,j}$. By these relations, we have the following diagram:

$$\begin{array}{ccccc}
\mathcal{B}_{m,n} & \xrightarrow{\cap H_{m,n}} & \mathcal{B}''_{m,n} & \xrightarrow{\cap (H_{m-1,n} \cap H_{m,n})} & (\mathcal{B}''_{m,n})'' \\
\downarrow{\scriptstyle \setminus H_{m,n}} & & \downarrow{\scriptstyle \setminus (H_{m-1,n} \cap H_{m,n})} & & \\
\mathcal{B}'_{m,n} & \xrightarrow{\cap H_{m-1,n}} & (\mathcal{B}'_{m,n})'' \text{ "} = \text{" } (\mathcal{B}''_{m,n})' & & \\
\downarrow{\scriptstyle \setminus H_{m-1,n}} & & & & \\
(\mathcal{B}'_{m,n})' & & & &
\end{array}$$

To extract a recurrence relation from this diagram, we introduce another notation: Let

$$\mathcal{B}^{\ell}_{m,n} = \{X_{i,j} \subset \mathbb{R}^{\ell} \mid i = 1, \ldots, m, j = 1, \ldots, n\}$$

be a hyperplane arrangement in $\mathbb{R}^{\ell}$ satisfying the following relations:

$$X_{i_1,j} \cap X_{i_2,j} \cap X_{i_3,j} = \emptyset, \tag{B.3}$$

$$X_{i_1,j_1} \cap X_{i_1,j_2} \cap X_{i_2,j_1} = X_{i_1,j_1} \cap X_{i_1,j_2} \cap X_{i_2,j_2} = X_{i_1,j_1} \cap X_{i_2,j_1} \cap X_{i_2,j_2} \tag{B.4}$$

for $i_1, i_2, i_3 = 1, \ldots, m$ and $j, j_1, j_2 = 1, \ldots, n$. Then, by the above arguments and a simple consideration, we have the following diagram:

$$
\begin{array}{ccccc}
\mathcal{B}^n_{m,n} & \xrightarrow{\cap X_{m,n}} & (\mathcal{B}^n_{m,n})'' & \xrightarrow{\cap (X_{m-1,n} \cap X_{m,n})} & \mathcal{B}^{n-2}_{m-1,n-1} \\
\downarrow{\scriptstyle \setminus X_{m,n}} & & \downarrow{\scriptstyle \setminus (X_{m-1,n} \cap X_{m,n})} & & \\
(\mathcal{B}^n_{m,n})' & \xrightarrow{\cap X_{m-1,n}} & ((\mathcal{B}^n_{m,n})')'' & \xrightarrow{\cap (X_{m-2,n} \cap X_{m-1,n})} & \mathcal{B}^{n-2}_{m-1,n-1} \\
\downarrow{\scriptstyle \setminus X_{m-1,n}} & & \downarrow{\scriptstyle \setminus (X_{m-2,n} \cap X_{m-1,n})} & & \\
\vdots & & \vdots & & \\
\downarrow{\scriptstyle \setminus X_{2,n}} & & \downarrow{\scriptstyle \setminus (X_{1,n} \cap X_{2,n})} & & \\
\mathcal{B} & \xrightarrow{\cap X_{1,n}} & \mathcal{B}^{n-1}_{m,n-1} & & \\
\downarrow{\scriptstyle \setminus X_{1,n}} & & & & \\
\mathcal{B}^n_{m,n-1} & & & &
\end{array}
\tag{B.5}
$$

Here, $\mathcal{B}$ is the hyperplane arrangement in $\mathbb{R}^n$ defined by

$$
\mathcal{B} = \mathcal{B}^n_{m,n-1} \cup \{X_{1,n}\}.
$$

Let $b^\ell_{m,n} = |\mathrm{Ch}(\mathcal{B}^\ell_{m,n})|$. Then, by Theorem 1 with the diagram (B.5), we have the recurrence relation

$$
b^n_{m,n} = b^n_{m,n-1} + m b^{n-1}_{m,n-1} + \frac{m(m-1)}{2} b^{n-2}_{m-1,n-1}.
$$

Moreover, by considering recursively, we can show that the following holds for $\ell, m, n \geq 1$:

$$
b^\ell_{m,n} = b^\ell_{m,n-1} + m b^{\ell-1}_{m,n-1} + \frac{m(m-1)}{2} b^{\ell-2}_{m-1,n-1}. \tag{B.6}
$$

Here, $b^0_{m,n} = b^\ell_{0,n} = b^\ell_{m,0} = 1$ for any $\ell, m, n \geq 0$ and we set $b^\ell_{m,n} = 0$ for $\ell < 0$. Then, for example, by (B.6), we have $b^1_{m,n} = mn + 1$ for any $m, n \geq 0$, $b^\ell_{m,1} = m^2/2 + m/2 + 1$ for any $\ell \geq 2$ and m. In particular, $b^\ell_{m,n}$ is a polynomial with respect to m.

By this recurrence relation (B.6), we can represent $b^n_{m,n}$ as

$$
b^n_{m,n} = \sum_{k=0}^{n/2} \sum_{\ell=0}^{n} d_{\ell,k}(m) b^{n-2k-\ell}_{m-k,0} = \sum_{k=0}^{n/2} \sum_{\ell=0}^{n} d_{\ell,k}(m),
$$

where $d_{\ell,k}(m)$ is a non-negative integer. Here, the last equation follows from $b^{n-2k-\ell}_{m-k,0} = 1$ for any k, ℓ, m such that $m - k \geq 0$ and $n - 2k - \ell \geq 0$. Then, it is easy to show that $d_{\ell,k}(m)$ is obtained as a sum of multiples of k times "$m(m-1)/2$", ℓ times "m", and $n - k - \ell$ times 1. Here, these double quotation means that these vary in accordance with the order of the operations. Indeed, the iteration relation (B.6) can be represented as a higher-dimensional analogue of Pascal's triangle as Figure 5. However, because we will calculate only the coefficient of leading term of $b^n_{m,n}$ as a polynomial of variable m, we may not take care of the orders. Then, the degree of $d_{\ell,k}(m)$ as a polynomial of variable m is equal to $2k + \ell$. This means that the leading term of $b^n_{m,n}$ as a polynomial of variable m is equal to the sum of terms $d_{\ell,k}(m)$ for $2k + \ell = n$. Moreover, by the fact $d_{\ell,k}(m) \geq 0$, we have

$$
b^n_{m,n} = \sum_{k=0}^{n/2} \sum_{\ell=0}^{n} d_{\ell,k}(m) \geq \sum_{k=0}^{n/2} d_{n-2k,k}(m) = (\text{the leading term of } b^n_{m,n} \text{ as a polynomial of variable } m).
$$

We calculate a lower bound of the leading term. Then, the leading term of $d_{n-2k,k}(m)$ as a polynomial of m can be written as

$$
d_{n-2k,k}(m) = \binom{n}{k, k, n-2k} \frac{1}{2^k} m^n + O(m^{n-1}),
$$

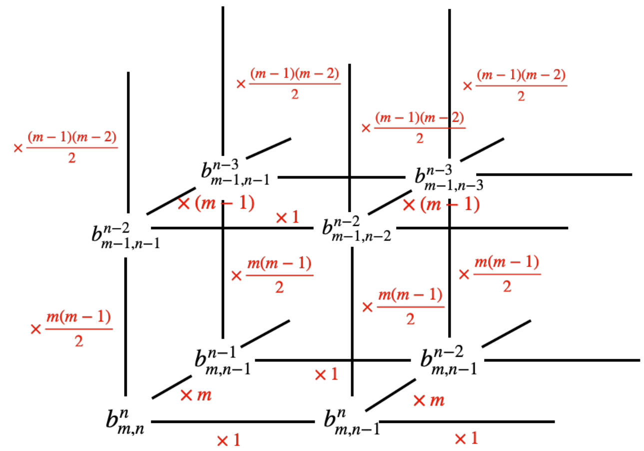

Figure 5: A higher dimensional analogue of Pascal's triangle representing the iteration relation (B.6).

where $\binom{n}{k_1,\ldots,k_m}$ for positive integers $k_1, \ldots, k_m$ such that $n = k_1 + \cdots + k_m$ is the multinomial coefficient defined by

$$\binom{n}{k_1,\ldots,k_m} = \frac{n!}{k_1!\cdots k_m!} = \binom{k_1}{k_1}\binom{k_1+k_2}{k_2}\cdots\binom{k_1+k_2+\cdots+k_m}{k_m}. \tag{B.7}$$

Indeed, as mentioned before, $d_{n-2k,k}(m)$ is obtained as a sum of multiples of k times of "$m(m-1)/2$", $n-2k$ times of "m", and k times of 1. Although the terms in the double quotations varies in accordance with the orders of the operations, the leading term is independent of the orders. Hence, the leading term of $d_{n-2k,k}(m)$ is the sum of multiples of k times of $1/2$, $n-2k$ times of 1, and k times of 1. The number of such multiples in the sum is same as $\binom{n}{k,k,n-2k}$. Hence, we have

$$d_{n-2k,k}(m) = \binom{n}{k,k,n-2k}\frac{1}{2^k}m^n + O(m^{n-1}).$$

By the form of RHS of equation (B.7) and the estimate in (2.4), we have

$$\begin{aligned}\binom{n}{k,k,n-2k} &= \binom{k}{k}\binom{2k}{k}\binom{n}{n-2k} = \binom{2k}{k}\binom{n}{n-2k}\\ &\geq \frac{2^{kH(1/2)}}{\sqrt{8k(1-1/2)}}\frac{2^{nH((n-2k)/n)}}{\sqrt{8k(n-2k)(1-(n-2k)/n)}}\\ &= \frac{2^{2k}2^{nH((n-2k)/n)}}{8k\sqrt{(n-2k)/n}}.\end{aligned}$$

In the last inequality follows from $H(1/2) = 1$.

We evaluate the coefficient of the leading term at $k = n/4$. Then, we have

$$d_{n/2,n/4}(m) \geq \frac{(2^{5/4})^n}{n\sqrt{2}}m^n + O(m^{n-1}).$$

In particular, the coefficient of leading term of $b^n_{m,n}$ is bounded from below by $(2^{5/4})^n/(n\sqrt{2})$. This concludes the proof. □

C Proof of Proposition 2

Proof of Proposition 2. Let $\lambda \in \Lambda$, $\boldsymbol{x} \in D_\lambda$ and $\phi \in \Phi$. We assume that ϕ satisfies (1) $\phi(D_\lambda) = D_{\lambda'}$ and (2) $f_\lambda = f_{\lambda'} \circ \phi|_{D_\lambda}$. Then, we have

$$\begin{aligned} f(\phi(\boldsymbol{x})) &= f_{\lambda'}(\phi(\boldsymbol{x})) = (f_{\lambda'} \circ \phi|_{D_\lambda})(\boldsymbol{x})\\ &= f_\lambda(\boldsymbol{x}) = f(\boldsymbol{x}). \end{aligned} \tag{C.1}$$

This equation holds for any $\boldsymbol{x}$ and any $\phi \in \Phi$. Because $\phi \in \Phi$ is a Euclidean transformation, ϕ is an isomorphism. In particular, the inverse of ϕ exists. As for any $\boldsymbol{y} \in \mathbb{R}^n$, there is a $\boldsymbol{x}$ such that $\boldsymbol{y} = \phi(\boldsymbol{x})$, by the equation (C.1), we have

$$f(\phi^{-1}(\boldsymbol{y})) = f(\boldsymbol{x}) = f(\phi(\boldsymbol{x})) = f(\boldsymbol{y}). \tag{C.2}$$

Hence, f is invariant by the action of ϕ^{-1} for any $\phi \in \Phi$. Now, let $\widehat{\Phi}$ be the subgroup of the group of Euclidean transformations generated by Φ. This means that any element $\phi \in \widehat{\Phi}$ is a composition of finite elements of $\{\phi_1, \dots, \phi_t, \phi_1^{-1}, \dots, \phi_t^{-1}\}$. Hence, by combining this fact and equations (C.1) and (C.2), f is invariant by the action of the group $\widehat{\Phi}$. $\square$

References

Orlik, P. and Terao, H. (2013). *Arrangements of hyperplanes*, volume 300. Springer Science & Business Media.