Subrepresentations

BY DENIS ROSSET

November 24, 2020

1 Definitions

Let $V = \mathbb{F}^D$ be a vector space where $\mathbb{F} = \mathbb{R}, \mathbb{C}$, and ρ a representation of a group G such that

$$\rho: G \to \mathrm{GL}(V). \tag{1}$$

While V is the Euclidean space, we consider its inner product structure as somehow arbitrary. We do not assume that ρ is unitary. Equivalently, we do not assume that the Euclidean inner product $(\cdot, \cdot)_E$ is ρ-invariant: i.e. $(\rho_g \vec{x}, \rho_g \vec{y})_E \neq (\vec{x}, \vec{y})_E$ in general. At any place we use the inner product, or the "conjugate transpose", any other inner product/transpose map can be substituted.

Definition 1. *For a map $X: V \to V$, not necessarily equivariant, we define $\Sigma_\rho[X] = \int_G \mathrm{d}\mu_g\, \rho_g X \rho_g^{-1}$. Consequently, $\Sigma_\rho[X] \rho_g = \rho_g \Sigma_\rho[X]$ is equivariant.*

We come to our main construction.

Proposition 2. *Let $W = \mathbb{F}^d$. Let (I, P) be a pair composed of the (injective)* injection map *$I: W \to V$ and* projection map *$P: V \to W$, with $\Pi \equiv IP$. If*

$$PI = \mathbb{1}, \qquad \text{and} \qquad \Pi \rho_g = \rho_g \Pi, \quad \forall g \in G, \tag{2}$$

then $\sigma_g \equiv P \rho_g I$ defines a representation $\sigma: G \to \mathrm{GL}(W)$ of G.

Proof. We have $\sigma_g \sigma_h = P \rho_g I P \rho_h I = P \rho_g \rho_h I P I = P \rho_{gh} I = \sigma_{gh}$. □

Note that our requirements imply that the $\Pi^2 = IPIP = IP = \Pi$ is a projector. Let $V_I = \mathrm{range}(I)$; observe that Π is a projector on V_I, and that V_I is a ρ-invariant subspace of V. Thus, (I, P) also defines a subrepresentation of ρ in V on the invariant subspace V_I.

Remark 3. If the representation ρ is unitary, we can assume that I is an isometry ($I^\dagger I = \mathbb{1}$) and take $P = I^\dagger$; but we do not assume this below. Our derivation is better understood keeping I and P distinct.

Note 4. In the "folklore", representations are usually defined using a basis $\{\vec{v}_1, \ldots, \vec{v}_d\}$. By writing $\vec{v}_i = I \vec{e}_i$ for the Euclidean basis vectors $\{\vec{e}_i\}$ of W, the basis is given by the columns of I written in matrix form. However, the corresponding projection P is not necessarily unique, nor the projector $\Pi = IP$. How come?

It turns out that we did not equip our space V with an inner product; in the textbook treatment, that is one of the first steps: construct a G-invariant inner product. However, that product is not necessarily be unique. Indeed, consider the case of a trivial group $G = \{e\}$ and a trivial representation ρ with $D = \dim V > 1$. Then any inner product on V is G-invariant. Now, consider a $d = \dim V_I = 1$ invariant subspace spanned by $\vec{v}_1$ with I fully specified by $I \vec{e}_1 = \vec{v}_1$. Then there are infinitely many maps $P: V \to W$ such that $P \vec{v}_1 = \vec{e}_1$.

2 Incomplete information

There are three ways to define a subrepresentation of $\rho: G \to \mathrm{GL}(V)$:

1. Provide a subspace $V_I \subseteq V$.
2. Consider a projector $\Pi: V \to V$ commuting with ρ.
3. Consider a pair of maps (I, P) satisfying the requirements of Proposition 2.

Note that $V_I = \operatorname{range} \Pi$, and $\Pi = IP$, and thus the quantity of information increases from 1. to 3.

Conversely, in 2. the subrepresentation σ_g is defined only up to a similarity transformation $\sigma_g \to T\sigma_g T^{-1}$. In 1., the projector Π is not uniquely defined in case of nontrivial multiplicities.

2.1 Reconstructing missing information

To recover Π from V_I, we have the following proposition.

Proposition 5. *Let $V_I \subseteq V$ be a ρ-invariant subspace. Then the map $\Pi: V \to V$*

$$\Pi = \Sigma_\rho[I(I^\dagger I)^{-1} I^\dagger] \tag{3}$$

is a projector with range V_I which commutes with ρ. We use an arbitrary inner product to compute a conjugate transpose $I^\dagger: V \to W$.

Proof. We verify that $I^\dagger$ is surjective and that $I^\dagger I$ is not singular. Then $\hat{\Pi} = I(I^\dagger I)^{-1} I^\dagger$ has range V_I by construction and is a projector:

$$\hat{\Pi}^2 = I(I^\dagger I)^{-1} I^\dagger I (I^\dagger I)^{-1} I^\dagger = \hat{\Pi}. \tag{4}$$

Now $\hat{\Pi}$ does not commute with ρ, but $\Pi = \Sigma_\rho[\hat{\Pi}]$ does, and has for range V_I (prove this last thing!). $\square$

To reconstruct a pair (I, P) from a projector Π, we have the following proposition.

Proposition 6. *Let $\Pi: V \to V$ be a G-equivariant projector. Then there exists a pair (I, P) that obeys the requirements of Proposition 2 such that $\Pi = IP$.*

Proof. Let $W = \mathbb{F}^d$ with d the rank of Π. Construct the injection map $I: W \to V$ using either:

- a set of d linearly independent columns of the matrix defining $\Pi: \mathbb{F}^D \to \mathbb{F}^D$,
- sampling a random map $\hat{I}: W \to V$ such that $I = \Pi \hat{I}$ has range V_I.

Then there exists a unique P such that $\Pi = IP$. $\square$

For fun, consider Maschke's theorem.

Proposition 7. *Let V_I be a G-invariant subspace of V. Then there exists a G-invariant subspace $V_{I'}$ such that $V = V_I \oplus V_{I'}$.*

Proof. Construct Π with range V_I using Proposition 5. Let $\Pi' = \mathbb{1} - \Pi$. Then $\Pi'\Pi = \Pi\Pi' = 0$ and thus $V_{I'} = \operatorname{range}\Pi'$ is a complement of V_I. $\square$

3 Approximate subrepresentations

We will need the following.

Definition 8. *The* condition number C_ρ *of a representation* ρ *is given by*

$$C_\rho = \min_T \operatorname{cond}(T), \qquad \textit{s.t. } T\rho T^{-1} \textit{ is unitary.} \tag{5}$$

It is also an upper bound on the operator norm $\|\rho_g\|_2$.

3.1 Computing the error affecting a subrepresentation

In general, we do not have access to an exact pair (I, P), rather to an approximation $(\tilde{I}, \tilde{P})$. We shall neglect the floating point errors in the relation $\tilde{P} \cdot \tilde{I} = \mathbb{1}$. We assume that the main source of error is that the relation $\tilde{\Pi}\rho_g = \rho_g\tilde{\Pi}$, with $\tilde{\Pi} = \tilde{I}\,\tilde{P}$ is satisfied only approximately.

To construct an error model for $(\tilde{I}, \tilde{P})$, we write $\tilde{I} = I + \Delta_I$ and $\tilde{P} = P + \Delta_P$ where (I, P) satisfies the conditions of Proposition 2. Thus, I and $\tilde{I}$ have the same dimensions, and our construction works only if an invariant subspace V_I of dimension $d = \operatorname{rank}(\tilde{I})$ exists.

We assume $\Delta_P \cdot I = 0$ (or $\tilde{P} \cdot I = \mathbb{1}$); otherwise, we write $A = \tilde{P} \cdot I$, and set $I \rightarrow I \cdot A^{-1}$ and $P \rightarrow A \cdot P$.

We write the exact representation $\sigma_g = P\rho_g I$ and the approximate representation $\tilde{\sigma}_g = \tilde{P}\rho_g\tilde{I}$.

Then, we measure the error by the Frobenius norms of $\|P\Delta_I\|_F$, $\|\Delta_P\|_F$ and $\|\Delta_I\|_F$. We compute:

$$\varepsilon = \|\tilde{\sigma}_g - \sigma_g\|_F = \|\tilde{P}\rho_g\tilde{I} - P\rho_g I\|_F = \|(P + \Delta_P)\rho_g(I + \Delta_I) - P\rho_g I\|_F = \|\Delta_P\,\rho_g\,I + P\rho_g\,\Delta_I + \Delta_P\,\rho_g\,\Delta_I\|_F. \tag{6}$$

Now, $\Delta_P\rho_g I = \Delta_P I\sigma_g = 0$. Then

$$\varepsilon = \|P\rho_g\Delta_I + \Delta_P\rho_g\Delta_I\|_F \leqslant \|\sigma_g P\Delta_I\|_F + \|\Delta_P\|_F \cdot \|\rho_g\Delta_I\|_2 \leqslant \|\sigma_g\|_2 \cdot \|P\Delta_I\|_F + \|\Delta_P\|_F \cdot \|\Delta_I\|_F \cdot \|\rho_g\|_2, \tag{7}$$

and $\varepsilon \leqslant C_\sigma \|P\,\Delta_I\|_F + \|\Delta_P\|_F \cdot \|\Delta_I\|_F \cdot C_\rho$.

Finally, $C_\sigma \leqslant \operatorname{cond}(P) \cdot C_\rho$, where $\operatorname{cond}(P) = \operatorname{cond}(I) \cong \operatorname{cond}(\tilde{P}) \cong \operatorname{cond}(\tilde{I})$ is the condition number of the matrices defining (I, P) and $(\tilde{I}, \tilde{P})$.

3.2 Finding a pair $(\boldsymbol{I}, \boldsymbol{P})$ close to an approximate pair $(\tilde{\boldsymbol{I}}, \tilde{\boldsymbol{P}})$

Given $(\tilde{I}, \tilde{P})$, we set $\tilde{P}_1 = \tilde{P}$ and $\tilde{I}_1 = \tilde{I}$. We then perform the iterations below, starting with $k = 1$. In the notation below, $V = V_I \oplus V_{I'} \oplus V_0$ is decomposed into invariant subspaces V_I, $V_{I'}$ and V_0.

We assume that V_I is the invariant subspace we are looking for; that V_0 is an invariant subspace containing additional multiplicity spaces for the irreducible representations present in V_I (we can always choose V_I and V_0 such that $\tilde{P}$ does not have support in V_0); that $V_{I'}$ is an invariant subspace whose irreducible representations do not overlap with V_I.

1. We decompose $\tilde{I}_k = I_k + \Delta_{I_k}$, for unknown I_k and Δ_{I_k} with $\operatorname{range} I_k = V_I$ and $\operatorname{range} \Delta I_k = V_{I'}$.

2. We decompose $\tilde{P}_k = P_k + \Delta_{P_k}$, for unknown P_k and Δ_{P_k} such that $P_k \cdot \Delta I_k = 0$ and $\Delta P_k \cdot I_k = 0$.

(We do not necessarily have $P_k I_k = \mathbb{1}$.)

3. We compute $\tilde{F}_k = \tilde{I}_k \tilde{P}_k = I_k P_k + I_k \Delta_{P_k} + \Delta_{I_k} P_k + \Delta_{I_k} \Delta_{P_k}$.

4. We compute $\overline{F_k} = \Sigma_\rho(\tilde{F}_k) = I_k P_k + \Delta_{I_k} \Delta_{P_k}$ (by Schur's lemma).

5. We monitor the norm $\|\tilde{F}_k - \overline{F_k}\|_F$, which should decrease until machine precision is reached.

6. We compute, by solving a linear least-squares problem:

$$
\begin{aligned}
I_k' &= \operatorname*{argmin}_{I_k'} \|\overline{F_k} - I_k' \tilde{P}_k\|_F \\
&= \operatorname*{argmin}_{I_k'} \|I_k P_k + \Delta_{I_k} \Delta_{P_k} - I_k'(P_k + \Delta_{P_k})\|_F \\
&= \operatorname*{argmin}_{I_k'} \|(I_k - I_k') P_k + (\Delta_{I_k} - I_k') \Delta_{P_k}\|_F
\end{aligned}
$$

   and if $P_k \gg \Delta P_k$, then $I_k' \approx I_k$.

7. Similarly, we compute $P_k' = \operatorname*{argmin}_{P_k'} \|\overline{F_k} - \tilde{I}_k P_k'\|_F$, and $P_k' \approx P_k$.

8. We set $\tilde{I}_{k+1} = I_k' (\tilde{P} \cdot I_k')^{-1}$ to have $\tilde{P} \cdot \tilde{I}_{k+1} = \mathbb{1}$, where $\tilde{P}$ is the original approximate projection.

9. We set $\tilde{P}_{k+1} = (P_k' \cdot \tilde{I}_{k+1}) P_k'$ to have $\tilde{P}_{k+1} \tilde{I}_{k+1} = \mathbb{1}$.

10. We monitor $\|\tilde{I}_{k+1} - \tilde{I}_k\|_F$ and $\|\tilde{P}_{k+1} - \tilde{P}_k\|_F$ to stop the iterations. (Stopping criterion to be fixed)

Question: can we repeat steps 6-7 without recomputing $\overline{F_k}$? Use iterative refinement?

3.3 A faster algorithm, unitary version

Let $\tilde{Q}_1: W \to V$ be an isometry whose range approximates a subspace of V invariant under ρ, and ρ be unitary. Let $K \geqslant 0$ be an integer. One step of the refinement process works as follows.

1. For the step k:

2. If $K = 0$, compute $\overline{F_k} = \Sigma_\rho(\tilde{Q}_k \tilde{Q}_k^\dagger)$. If $K \geqslant 1$, compute $\overline{F_k} = (\sum_{i=1}^K \rho_{g_i} \tilde{Q}_k \tilde{Q}_k^\dagger \rho_{g_i}^{-1}) / K$, where $\{g_i\}$ are sampled uniformly randomly from G.

3. Let X minimize $\| \tilde{Q}_k Y_k^\dagger - \overline{F_k}\|_2$. The solution is $Y_k^\dagger = \tilde{Q}_k^\dagger \overline{F_k}$. Set $\tilde{Q}_{k+1}$ to an isometry that has the same range as Y_k (using one of Gram-Schmidt, QR decomposition, SVD).

4. Either repeat steps 1. and 2.; or repeat only step 2 a few times by incrementing k but reusing $\overline{F_{k+1}} = \overline{F_k}$.

Finally, remark that the product $Y_k = \overline{F_k} \tilde{Q}_k$ can be computed faster, for $k \geqslant 1$, as:

$$
Y_k = \sum_i \rho_{g_i} \tilde{Q}_k \tilde{Q}_k^\dagger \rho_{g_i}^{-1} \tilde{Q}_k = \sum_i \underbrace{(\rho_{g_i} \tilde{Q}_k)}_{n \times m} \underbrace{(\tilde{Q}_k^\dagger \rho_{g_i}^{-1} \tilde{Q}_k)}_{m \times m}, \tag{8}
$$

assuming that we have a fast method to compute products such as $A \rho_g B$ and $\rho_g B$. This is the case for permutation representations, tensor products, etc. In that case, we perform step 2. only once (there is no point in reusing the same samples).

Note 9. Now, we prove that the solution of $\min \|AX-B\|_F$ is $X = A^\dagger B$ if A is an isometry ($A^\dagger A = \mathbb{1}$). Let $A^\perp$ be the matrix orthogonal to A such that the concatenation $(A, A^\perp)$ is a unitary matrix.

Then

$$\|AX-B\|_F = \left\|(A,A^\perp)\begin{pmatrix} X \\ 0 \end{pmatrix} - B\right\|_F = \left\|(A,A^\perp)\begin{pmatrix} X \\ 0 \end{pmatrix} - (A,A^\perp)\begin{pmatrix} A^\dagger \\ (A^\perp)^\dagger \end{pmatrix} B\right\|_F \tag{9}$$

Now, as the Frobenius norm is invariant under unitary transformations:

$$\min \|AX-B\|_F = \left\|\begin{pmatrix} X \\ 0 \end{pmatrix} - \begin{pmatrix} A^\dagger \\ (A^\perp)^\dagger \end{pmatrix} B\right\|_F = \|X - A^\dagger B\|_F + \|(A^\perp)^\dagger B\|_F. \tag{10}$$

Thus $X = A^\dagger B$.

3.4 Faster algorithm, non unitary version

Let $(\tilde{I}, \tilde{P})$ be an approximate pair.

1. Set $I_1 = \tilde{I}$,$P_1 = \tilde{P}$. Set $k \leftarrow 1$.
2. Compute temporaries $J_k = \sum_{i=1}^K (\rho_{g_i} I_k)(P_k \rho_{g_i}^{-1} I_k)$ and $Q_k = \sum_{i=1}^K (P_k \rho_{g_i} I_k)(\rho_{g_i}^{-1} P_k)$.
3. Set $I_{k+1} = J_k(\tilde{P} J_k)^{-1}$, to recover $\Delta P \cdot I_{k+1} = 0$.
4. Set $P_{k+1} = (Q_k I_{k+1})^{-1} Q_k$ to recover $P_{k+1} I_{k+1} = \mathbb{1}$.
5. $k \leftarrow k+1$ and loop.

3.5 Error estimation

We cannot check $\|\tilde{F}_k - \overline{F_k}\|_F$ anymore as we only approximate it. However, we can check how the subspaces spanned by $\tilde{Q}_k$ and $\tilde{Q}_{k+1}$ are aligned. We construct the matrix Λ:

$$\Lambda = \tilde{Q}_k^\dagger \tilde{Q}_{k+1}. \tag{11}$$

Then, if $\tilde{Q}_k$ and $\tilde{Q}_{k+1}$ span the same space, they will be related by a unitary transformation, and the singular values of Λ will all be one. We thus take

$$\delta = 2\|\tilde{Q}_k^\dagger \tilde{Q}_{k+1} \tilde{Q}_{k+1}^\dagger \tilde{Q}_k - \mathbb{1}\|_F \tag{12}$$

to compute $\sqrt{\sum_i (\lambda_i - 1)^2}$, where the λ_1 are the singular values of Λ (TODO: prove this, and why is there a factor 2).