A PROOF

A.1 BUILDING THE HARD FUNCTION G

For any $r \in \mathbb{R}$ with $|r| < 1$, by mapping $x \mapsto rx$ and using homogeneity of s_λ and V, we define G via the generating function identity:

$$G := \frac{C}{\sqrt{N!}} \sum_{\lambda \text{ doubly even}} r^{\left(|\lambda| + \frac{N(N-1)}{2}\right)} s_\lambda \cdot V = C\sqrt{N!} \cdot \prod_{i<j} \frac{1}{1 - r^4 x_i^2 x_j^2} \cdot \mathcal{A}(\phi_1^{(r)} \otimes \cdots \otimes \phi_N^{(r)}) \ , \tag{21}$$

where

$$\phi_j^{(r)}(x_i) = \begin{cases} rx_i((rx_i)^2)^{N/2-j}(1+(rx_i)^4)^{j-1} & \text{if } 1 \leq j \leq N/2 \\ ((rx_i)^2)^{N-j}(1+(rx_i)^4)^{j-1-N/2} & \text{if } N/2+1 \leq j \leq N \end{cases} \tag{22}$$

where C is chosen to normalize G.

Note that from the RHS, it is clear that G is written in the form of a Jastrow ansatz. We will discuss efficiency of computing G further below.

It remains to choose r and C such that $\|G\| = 1$. Note that, if $p(k)$ denotes the number of partitions of k, and and $p'(k)$ denotes the number of doubly even partitions of k , it's easy to see that

$$p'(k) = \begin{cases} p(k/4) & k \equiv 0 \mod 4 \\ 0 & else \end{cases} \tag{23}$$

So we calculate by orthogonality:

$$\begin{aligned} \|G\|^2 &= \frac{C^2}{N!} \left\langle \sum_{\lambda \text{ doubly even}} r^{\left(|\lambda| + \frac{N(N-1)}{2}\right)} s_\lambda \cdot V, \sum_{\mu \text{ doubly even}} r^{\left(|\mu| + \frac{N(N-1)}{2}\right)} s_\mu \cdot V \right\rangle & (24) \\ &= C^2 r^{N(N-1)} \sum_{\lambda \text{ doubly even}} r^{2|\lambda|} & (25) \\ &= C^2 r^{N(N-1)} \sum_{k=0}^{\infty} r^{2k} p'(k) & (26) \\ &= C^2 r^{N(N-1)} \sum_{k=0}^{\infty} r^{8k} p(k) & (27) \\ &= C^2 r^{N(N-1)} \prod_{k=1}^{\infty} \frac{1}{1 - r^{8k}} & (28) \end{aligned}$$

where in the last line we employ the generating function for partition numbers. Then setting $C = \left(r^{-N(N-1)} \prod_{k=1}^{\infty} 1 - r^{8k}\right)^{1/2}$ gives $\|G\| = 1$.

A.2 FROM TENSORS TO MATRICES

The point of choosing G in this way, is it enables a simple flattening argument, where we can reduce comparing tensors to comparing matrices.

Note again that terms of the form x^α for $\alpha \in \mathbb{N}^N$ are orthonormal. Hence, we derive an initial lower bound by Bessel's inequality:

$$\|F - G\|^2 \geq \sum_{\alpha \in \mathbb{N}^N} \left(\langle F, x^\alpha \rangle - \langle G, x^\alpha \rangle\right)^2 \ . \tag{29}$$

Note that by antisymmetry of F and G, if α doesn't have distinct elements then

$$\langle F, x^\alpha \rangle = \langle G, x^\alpha \rangle = 0 \, . \tag{30}$$

To see this, suppose $\alpha_1 = \alpha_2$, and let P_{12} be the permutation operator defined by

$$P_{12} F(x_1, x_2, x_3, \dots) = F(x_2, x_1, x_3, \dots) \tag{31}$$

It's easy to see P_{12} is a symmetric operator with respect to $\langle \cdot, \cdot \rangle$. Then for any antisymmetric function H,

$$\begin{aligned} \langle H, x^\alpha \rangle &= \langle H, P_{12} x^\alpha \rangle & (32) \\ &= \langle P_{12} H, x^\alpha \rangle & (33) \\ &= -\langle H, x^\alpha \rangle & (34) \end{aligned}$$

which implies $\langle H, x^\alpha \rangle = 0$.

Furthermore, let us define the equivalence class $\sim$ as via $\alpha \sim \alpha'$ if there exists a permutation π such that $\alpha = \pi \circ \alpha'$. Then by similar reasoning, $\alpha \sim \alpha'$ implies:

$$\langle F, x^\alpha \rangle = (-1)^\pi \left\langle F, x^{\alpha'} \right\rangle \tag{35}$$

$$\langle G, x^\alpha \rangle = (-1)^\pi \left\langle G, x^{\alpha'} \right\rangle \tag{36}$$

So define $\mathbb{N}^N_{\geq}$ to be the set of strictly decreasing non-negative integer vectors of length N, then we have:

$$\|F - G\|^2 \geq N! \cdot \sum_{\alpha \in \mathbb{N}^N_{\geq}} \left(\langle F, x^\alpha \rangle - \langle G, x^\alpha \rangle \right)^2 \tag{37}$$

Now, we can consider a flattening argument, by passing from tensors to matrices. Define

$$\mathfrak{A}_1 = \{ \beta \in \mathbb{N}^{N/2}_{\geq} : \beta_i \equiv 1 \mod 2 \} \tag{38}$$

$$\mathfrak{A}_2 = \{ \gamma \in \mathbb{N}^{N/2}_{\geq} : \gamma_i \equiv 0 \mod 2 \} \tag{39}$$

For $\beta \in \mathfrak{A}_1$ and $\gamma \in \mathfrak{A}_2$, let $\beta \cup \gamma \in \mathbb{N}^N$ be the concatenation of β and γ.

Then given a function acting on N particles such as G, we can map G to a (infinite-dimensional) matrix M indexed by the sets $\mathfrak{A}_1$ and $\mathfrak{A}_2$:

$$M(G) = \left[\langle G, x^{\beta \cup \gamma} \rangle \right]_{\beta, \gamma} \tag{40}$$

Let us calculate the entries of this matrix. Let $\delta = (N-1, N-2, \dots, 1, 0)$, and observe that:

$$\langle s_\lambda \cdot V, x^{\beta \cup \gamma} \rangle = \begin{cases} \pm 1 & \lambda + \delta \sim \beta \cup \gamma \, , \\ 0 & \text{otherwise.} \end{cases} \tag{41}$$

Note that ambiguity in sign depends on the sign of the permutation that maps $\lambda + \delta$ to $\beta \cup \gamma$.

By definition, G is a sum of terms of the form $s_\lambda \cdot V$ where λ is doubly even. This implies that $\lambda + \delta = (2a_1 + 1, 2a_1, 2a_2 + 1, 2a_2, \dots)$ with $a_i > a_{i+1}$. In other words, $\lambda + \delta \sim (\gamma + \mathbf{1}) \cup \gamma$ with $\gamma + \mathbf{1} \in \mathfrak{A}_1$ and $\gamma \in \mathfrak{A}_2$, where $\mathbf{1}$ is the all-ones vector. See Figure 2 for an example.

It follows that we may write:

$$\langle G, x^{\beta \cup \gamma} \rangle = \begin{cases} \pm \frac{C}{\sqrt{N!}} r^{\left(|\lambda| + \frac{N(N-1)}{2}\right)} & \beta = (\gamma + \mathbf{1}), \quad \lambda + \delta \sim (\gamma + \mathbf{1}) \cup \gamma \, , \\ 0 & \text{otherwise.} \end{cases} \tag{42}$$

Suppose we index $M(G)$ such that the ith column is indexed by $\gamma^{(i)}$, and the ith row is indexed by $\gamma^{(i)} + \mathbf{1}$. Then $M(G)$ is in fact a diagonal matrix. And given the functional form of G, we have that the diagonal terms will include:

- $\pm\frac{C}{\sqrt{N!}}r^{\left(0+\frac{N(N-1)}{2}\right)}$ repeated $p(0)$ times,
- $\pm\frac{C}{\sqrt{N!}}r^{\left(4+\frac{N(N-1)}{2}\right)}$ repeated $p(1)$ times,
- $\pm\frac{C}{\sqrt{N!}}r^{\left(8+\frac{N(N-1)}{2}\right)}$ repeated $p(2)$ times,
- $\dots$
- $\pm\frac{C}{\sqrt{N!}}r^{\left(4k+\frac{N(N-1)}{2}\right)}$ repeated $p(k)$ times.

Second, let us consider $M(f_1 \otimes \dots \otimes f_N)$. We can calculate the inner product of a rank-one function as the product of orbital inner products:

$$\langle f_1 \otimes \dots \otimes f_N, x^{\beta\cup\gamma}\rangle = \prod_{n=1}^{N/2}\langle f_n, y^{\beta_n}\rangle \prod_{n=1}^{N/2}\langle f_{N/2+n}, y^{\gamma_n}\rangle \tag{43}$$

where we introduce $y \in \mathbb{C}$ as a one-dimensional dummy variable for integration.

Define vectors $u \in \mathbb{C}^{|\mathfrak{A}_1|}$ and $v \in \mathbb{C}^{|\mathfrak{A}_2|}$ such that

$$u_\beta = \prod_{n=1}^{N/2}\langle f_n, y^{\beta_n}\rangle \;, \tag{44}$$

$$v_\gamma = \prod_{n=1}^{N/2}\langle f_{N/2+n}, y^{\gamma_n}\rangle \;. \tag{45}$$

Then it's clear that $M(f_1 \otimes \dots \otimes f_N) = uv^T$, i.e. it is rank-one. Consequently, because F is the sum of $L \cdot N!$ rank-one tensors, $M(F)$ will be rank at most $L \cdot N!$.

So we finally pass from tensors to matrices, and lower bound via the Frobenius norm $\|\cdot\|_F$:

$$\|F - G\|^2 \geq N! \cdot \sum_{\alpha \in \mathbb{N}^N_{\geq}} \left(\langle F, x^\alpha\rangle - \langle G, x^\alpha\rangle\right)^2 \tag{46}$$

$$\geq N! \cdot \|M(F) - M(G)\|_F^2 \tag{47}$$

Thus, because $M(F)$ is low-rank and $M(G)$ is chosen to be diagonal, we have an approachable infinite-dimensional matrix low-rank approximation problem.

A.3 Deriving the bound

By SVD, the optimal choice for F is to produce a diagonal matrix $M(F)$ of rank $L \cdot N!$ with the maximal singular values of G along the diagonal. So it only remains to calculate these terms, and lower bound the approximation.

So suppose we choose $L \leq e^{N^2}$. Noting that $N^N \leq e^{N^2}/14$ for $N \geq 3$:

$$L \cdot N! \leq e^{N^2} N^N \tag{48}$$

$$\leq e^{2N^2}/14 \tag{49}$$

$$\leq p(N^4) \tag{50}$$

where the last line follows from Corollary 3.1 in Maróti (2003).

So clearly $L \leq e^{N^2}$ guarantees that $L \cdot N! \leq \sum_{k=0}^{N^4} p(k)$.

Thus, since $M(F)$ is constrained to have rank $\leq L \cdot N!$, it will be diagonal with $\leq \sum_{k=0}^{N^4} p(k)$ terms, so that:

$$\|F - G\|^2 \geq N! \cdot \|M(F) - M(G)\|_F^2 \tag{51}$$

$$\geq N! \cdot \sum_{k=N^4+1}^{\infty} \left(\pm \frac{C}{\sqrt{N!}} r^{\left(4k + \frac{N(N-1)}{2}\right)} \right)^2 p(k) \tag{52}$$

$$= C^2 r^{N(N-1)} \sum_{k=N^4+1}^{\infty} r^{8k} p(k) \tag{53}$$

$$= 1 - C^2 r^{N(N-1)} \sum_{k=0}^{N^4} r^{8k} p(k) \tag{54}$$

where the last line follows as C was chosen so that $C^2 r^{N(N-1)} \sum_{k=0}^{\infty} r^{8k} p(k) = 1$.

Note that

$$\sum_{k=0}^{N^4} r^{8k} p(k) \leq \prod_{k=1}^{N^4} \frac{1}{1 - r^{8k}} \tag{55}$$

as the LHS is the generating function for partitions λ with $|\lambda| \leq N^4$, and the RHS is the generating function for partitions with all parts $\leq N^4$, which clearly dominates the LHS termwise.

So plugging back in the definition of $C = \left(r^{-N(N-1)} \prod_{k=1}^{\infty} 1 - r^{8k} \right)^{1/2}$:

$$\|F - G\|^2 \geq 1 - C^2 r^{N(N-1)} \prod_{k=1}^{N^4} \frac{1}{1 - r^{8k}} \tag{56}$$

$$= 1 - \prod_{k=N^4+1}^{\infty} 1 - r^{8k} \,. \tag{57}$$

Finally, by choosing $r = 1 - \frac{1}{8N^4+8}$, we have:

$$1 - \prod_{k=N^4+1}^{\infty} 1 - r^{8k} \geq 1 - \left(1 - r^{8N^4+8}\right) \tag{58}$$

$$= \left(1 - \frac{1}{8N^4+8}\right)^{8N^4+8} \tag{59}$$

$$\geq \left(1 - \frac{1}{16}\right)^{16} \tag{60}$$

$$\geq \frac{3}{10} \,, \tag{61}$$

where we use that the limit $\left(1 - \frac{1}{n}\right)^n$ increases monotonically in n. Hence, we conclude:

$$\|F - G\|^2 \geq \frac{3}{10} \,. \tag{62}$$

A.4 Efficiency of representing G

We remind the representation, where with $r = 1 - \frac{1}{8N^4+8}$ we have:

$$G = C\sqrt{N!} \cdot \prod_{i<j} \frac{1}{1 - r^4 x_i^2 x_j^2} \cdot \mathcal{A}(\phi_1^{(r)} \otimes \cdots \otimes \phi_N^{(r)}) \tag{63}$$

with

$$\phi_j^{(r)}(x_i) = \begin{cases} rx_i((rx_i)^2)^{N/2-j}(1+(rx_i)^4)^{j-1} & 1 \le j \le N/2\ , \\ ((rx_i)^2)^{N-j}(1+(rx_i)^4)^{j-1-N/2} & N/2+1 \le j \le N\ . \end{cases} \tag{64}$$

So it remains to characterize some $\hat{G}$ in the Jastrow ansatz, parameterized with neural networks, that approximates G.

First, we make mention that, for very specific activations, this function G may be written exactly in the Jastrow ansatz.

After fixing the value of r, we can consider each $\phi_i := \phi_i^{(r)}$ as an activation function acting on a one-dimensional input. Consider also the "activation" $\psi(x_i, x_j) = \frac{1}{1-r^4x_i^2x_j^2}$. Then G can be clearly written exactly in the Jastrow ansatz (where the Jastrow term J is given as a symmetric network with product pooling) as

$$G = C\sqrt{N!} \cdot J(x) \cdot \mathcal{A}(\phi_1 \otimes \cdots \otimes \phi_N) \tag{65}$$

$$= C\sqrt{N!} \cdot \prod_{i<j} \psi(x_i, x_j) \cdot \mathcal{A}(\phi_1 \otimes \cdots \otimes \phi_N) \tag{66}$$

We know consider approximation error under a more typical choice of activation function. We will use the modReLU Arjovsky et al. (2016):

$$\sigma(z) = \begin{cases} 0 & |z| \le 1 \\ z - \frac{z}{|z|} & |z| \ge 1 \end{cases} \tag{67}$$

We consider first the Jastrow factor. We will approximate it in $\hat{G}$ using a Relational Network (Santoro et al., 2017) with multiplication pooling.

Note that on our domain $(S^1)^N$, we have $1/2 < \frac{1}{|1-r^4x_i^2x_j^2|} < 8N^4 + 8$. The Lipschitz norm of this function is also polynomially bounded, and therefore from Theorem 1 in Caragea et al. (2022) it is standard to approximate this function with a complex neural network g with width, depth, and weights all in $O(poly(N, 1/\epsilon))$, such that $\|g\|_\infty \le 8N^4 + 8 + \epsilon$ and

$$\left\| \frac{1}{1-r^4x_i^2x_j^2} - g(x_i, x_j) \right\|_\infty \le \epsilon\ . \tag{68}$$

Assuming $\epsilon < 1$, it follows from routine calculation that

$$\left\| \prod_{i<j} \frac{1}{1-r^4x_i^2x_j^2} - \prod_{i<j} g(x_i, x_j) \right\|_\infty \le N^{O(N)}\epsilon\ . \tag{69}$$

Consider second the antisymmetric factor. Following the row transforms given in the proof of Lemma 3.4 in Ishikawa et al. (2006), the antisymmetric term may be equivalently written as:

$$\mathcal{A}(\phi_1^{(r)} \otimes \cdots \otimes \phi_N^{(r)}) = \mathcal{A}(\psi_1^{(r)} \otimes \cdots \otimes \psi_N^{(r)}) \tag{70}$$

with

$$\psi_j^{(r)}(x_i) = \begin{cases} rx_i((rx_i)^2)^{N/2-j}(1+(rx_i)^{4(j-1)}) & 1 \le j \le N/2\ , \\ ((rx_i)^2)^{N-j}(1+(rx_i)^{4(j-1-N/2)}) & N/2+1 \le j \le N\ . \end{cases} \tag{71}$$

It's easy to confirm that restricted to S^1, $\|\psi^{(r)}\|_\infty \leq 2$, and $\mathrm{Lip}(\psi^{(r)}) = O(poly(N))$. Therefore, it is routine again by Theorem 1 in Caragea et al. (2022) to approximate any $\psi^{(r)}$ with a neural network $\hat{\psi}$ parameterized with width, depth and weights $O\left(poly(N, 1/\epsilon)\right)$ such that

$$\|\psi^{(r)} - \hat{\psi}\|_\infty \leq \epsilon \; . \tag{72}$$

.

We also clearly have that $\|\hat{\psi}\|_\infty \leq 2 + \epsilon$. Now, we calculate:

$$\begin{aligned}
\left\|\mathcal{A}(\psi_1^{(r)} \otimes \cdots \otimes \psi_N^{(r)}) - \mathcal{A}(\hat{\psi}_1 \otimes \cdots \otimes \hat{\psi}_N)\right\|_\infty &= \left\|\frac{1}{N!}\sum_\sigma (-1)^\sigma \left(\psi_{\sigma(1)}^{(r)} \otimes \cdots \otimes \psi_{\sigma(N)}^{(r)} - \hat{\psi}_{\sigma(1)} \otimes \cdots \otimes \hat{\psi}_{\sigma(N)}\right)\right\|_\infty & (73)\\
&\leq \frac{1}{N!}\sum_\sigma \left\|\left(\psi_{\sigma(1)}^{(r)} \otimes \cdots \otimes \psi_{\sigma(N)}^{(r)} - \hat{\psi}_{\sigma(1)} \otimes \cdots \otimes \hat{\psi}_{\sigma(N)}\right)\right\|_\infty & (74)\\
&< N 3^{N-1}\epsilon & (75)
\end{aligned}$$

Finally, we combine the Jastrow factor and antisymmetric component. Let

$$\hat{G}(x) = C\sqrt{N!}\prod_{i<j} g(x_i, x_j)\mathcal{A}(\hat{\psi}_1 \otimes \cdots \otimes \hat{\psi}_N)(x) \; . \tag{76}$$

Then we calculate:

$$\begin{aligned}
\|G - \hat{G}\|_\infty &= C\sqrt{N!}\left\|\prod_{i<j}\frac{1}{1 - r^4 x_i^2 x_j^2} \cdot \mathcal{A}(\phi_1^{(r)} \otimes \cdots \otimes \phi_N^{(r)}) - \prod_{i<j} g(x_i, x_j) \cdot \mathcal{A}(\hat{\psi}_1 \otimes \cdots \otimes \hat{\psi}_N)\right\|_\infty & (77)\\
&\leq N^{O(N)}\epsilon \; . & (78)
\end{aligned}$$

So it remains to rescale $\epsilon \to \frac{\epsilon}{N^{O(N)}}$, which yields the expected bounds on width. $\square$