Appendix A: Properties of the Gradient of the Dilatable Entropy DGF for Sequence-Form Strategy Spaces

LEMMA 1. *At any point* $\boldsymbol{x} \in \operatorname{relint} Q$*, the gradients* $\nabla\tilde{\varphi}(\boldsymbol{x})$ *and* $\nabla\varphi(\boldsymbol{x})$ *differ only along orthogonal directions to the affine hull* $\operatorname{aff} Q$ *of* Q*, that is,*

$$(\nabla\tilde{\varphi}(\boldsymbol{x}) - \nabla\varphi(\boldsymbol{x}))^\top(\boldsymbol{y} - \boldsymbol{z}) = 0 \qquad \forall \boldsymbol{y}, \boldsymbol{z} \in Q.$$

Proof. We will prove the theorem by structural induction on the structure of the sequence-form strategy space. Specifically, given any decision point $j \in \mathcal{J}$, we focus on the subtree composed of j and all of its descendants $j' \succeq j$. As an intermediate step, we being by showing that for all $j \in \mathcal{J}$

$$\sum_{j' \succeq j} \sum_{a' \in A_{j'}} \frac{\partial(\tilde{\varphi} - \varphi)}{\partial x_{j'a'}}(\boldsymbol{x}) y_{j'a'} = \gamma_j \, y_{p_j} \log x_{p_j}. \tag{32}$$

Since $\boldsymbol{y}$ is arbitrary, the equality above will immediately imply that for any j such that $p_j = \varnothing$

$$\sum_{j' \succeq j} \sum_{a' \in A_{j'}} \frac{\partial(\tilde{\varphi} - \varphi)}{\partial x_{j'a'}}(\boldsymbol{x})(y_{j'a'} - z_{j'a'}) = \gamma_j (y_\varnothing - z_\varnothing) \log x_\varnothing = 0, \tag{33}$$

where the last equality follows from the fact that $y_\varnothing = z_\varnothing = 1$ since $\boldsymbol{y}, \boldsymbol{z}$ are sequence-form strategies. Now, since the set of sequences can be partitioned as $\Sigma = \{\varnothing\} \sqcup \bigsqcup_{j \in \mathcal{J}: p_j = \varnothing} \{j'a' : j' \succeq j, a' \in A_{j'}\}$, and furthermore $y_\varnothing = z_\varnothing = 1$, it is evident that (33) considered for the set $\{j \in \mathcal{J} : p_j = \varnothing\}$ immediately implies the statement.

In order to prove (32) we will use structural induction over the structure of the sequence-form decision problem.

- **Base case:** terminal decision points j. Consider a terminal decision point j, that is one for which $\mathcal{C}_{ja} = \emptyset$ for all $a \in A_j$. Then, for all $a \in A_j$ we obtain from direct inspection that

  $$\frac{\partial\tilde{\varphi}}{\partial x_{ja}}(\boldsymbol{x}) = w_{ja}(1 + \log x_{ja}); \qquad \frac{\partial\varphi}{\partial x_{ja}}(\boldsymbol{x}) = \gamma_j\left(1 + \log\frac{x_{ja}}{x_{p_j}}\right).$$

  Hence,

  $$\begin{aligned}\sum_{a \in A_j} \frac{\partial(\tilde{\varphi} - \varphi)}{\partial x_{ja}}(\boldsymbol{x}) y_{ja} &= \sum_{a \in A_j} \big(w_{ja} - \gamma_j + (w_{ja} - \gamma_j)\log x_{ja} + \gamma_j \log x_{p_j}\big) y_{ja} \\ &= \gamma_j \, y_{p_j} \log x_{p_j},\end{aligned}$$

  where we used the fact that by definition $\gamma_j = w_{ja} = 1$, as well as the fact that $\boldsymbol{y}$ is a sequence-form strategy and therefore $\sum_{a \in A_j} y_{ja} = y_{p_j}$. Since the only $j' \succeq j$ is j itself, the proof of the base case is complete.
- **Inductive step.** Consider a nonterminal j. Again by direct inspection, we obtain

  $$\begin{aligned}\frac{\partial\tilde{\varphi}}{\partial x_{ja}}(\boldsymbol{x}) &= w_{ja}(1 + \log x_{ja}) + \sum_{j' \in \mathcal{C}_{ja}} \gamma_{j'} \log |A_{j'}|; \text{ and} \\ \frac{\partial\varphi}{\partial x_{ja}}(\boldsymbol{x}) &= \gamma_j\left(1 + \log\frac{x_{ja}}{x_{p_j}}\right) - \sum_{j' \in \mathcal{C}_{ja}} \gamma_{j'} + \sum_{j' \in \mathcal{C}_{ja}} \gamma_{j'} \log |A_{j'}| \\ &= w_{ja} + \gamma_j \log x_{ja} - \gamma_j \log x_{p_j} + \sum_{j' \in \mathcal{C}_{ja}} \gamma_{j'} \log |A_{j'}|,\end{aligned}$$

where the last equality uses the definition of $w_{ja} := \gamma_j - \sum_{j' \in A_j} \gamma_{j'}$.

Now, the set of descendant decision points $\{j' \in \mathcal{J} : j' \succeq j\}$ can be partitioned as

$$\{j\} \sqcup \bigsqcup_{a \in A_j} \bigsqcup_{j' \in \mathcal{C}_{ja}} \{j'' \in \mathcal{J} : j'' \succeq j'\}.$$

Correspondingly, using the inductive hypothesis we have

$$\begin{aligned}
\sum_{j' \succeq j} \sum_{a' \in A_{j'}} \frac{\partial(\tilde{\varphi} - \varphi)}{\partial x_{j'a'}}(\boldsymbol{x}) y_{j'a'} &= \left(\sum_{a \in A_j} \frac{\partial(\tilde{\varphi} - \varphi)}{\partial x_{ja}}(\boldsymbol{x}) y_{ja} \right) + \sum_{a \in A_j} \sum_{j' \in \mathcal{C}_{ja}} \gamma_{j'}\, y_{ja} \log x_{ja} \\
&= \left(\sum_{a \in A_j} \Big((w_{ja} - \gamma_j) \log x_{ja} + \gamma_j \log x_{p_j} \Big) y_{ja} \right) + \sum_{a \in A_j} \sum_{j' \in \mathcal{C}_{ja}} \gamma_{j'}\, y_{ja} \log x_{ja} \\
&= \sum_{a \in A_j} \left(\left(w_{ja} - \gamma_j + \sum_{j' \in \mathcal{C}_{ja}} \gamma_{j'} \right) \log x_{ja} + \gamma_j \log x_{p_j} \right) y_{ja} \\
&= \sum_{a \in A_j} \gamma_j\, y_{ja} \log x_{p_j} = \gamma_j\, y_{p_j} \log x_{p_j},
\end{aligned}$$

where we again used the definition of $w_{ja} := \gamma_j - \sum_{j' \in A_j} \gamma_{j'}$, as well as the fact that $\boldsymbol{y}$ is a sequence-form strategy and therefore $\sum_{a \in A_j} y_{ja} = y_{p_j}$.

This concludes the induction proof. □

Appendix B: Technical Results about Scaled Extension

In what follows, we let aff S denote the affine hull of set S, that is, the set

$$\operatorname{aff} S := \left\{ \sum_{i=1}^{k} \alpha_i \boldsymbol{s}_i \;\middle|\; k \in \mathbb{N}_{>0}, \alpha_i \in \mathbb{R}, \boldsymbol{s}_i \in S, \sum_{i=1}^{k} \alpha_i = 1 \right\}.$$

LEMMA 4. *Let $\mathcal{U} \subset \mathbb{R}^m, \mathcal{V} \subset \mathbb{R}^n$ and $h : \mathbb{R}^m \to \mathbb{R}$ be an affine function, nonnegative on $\mathcal{U}$. Then,*

$$\operatorname{aff}(\mathcal{U} \overset{h}{\lhd} \mathcal{V}) = (\operatorname{aff} \mathcal{U}) \overset{h}{\lhd} (\operatorname{aff} \mathcal{V}).$$

Proof. We prove the two directions of inclusion separately.

($\subseteq$) Let $(\boldsymbol{u}, \boldsymbol{w}) \in \operatorname{aff}(\mathcal{U} \lhd^h \mathcal{V})$. We argue that $(\boldsymbol{u}, \boldsymbol{w}) \in (\operatorname{aff} \mathcal{U}) \lhd^h (\operatorname{aff} \mathcal{V})$. By definition of affine hull, there must exist $k \in \mathbb{N}_{>0}$, points $(\boldsymbol{u}_i, \boldsymbol{w}_i) = (\boldsymbol{u}_i, h(\boldsymbol{u}_i)\boldsymbol{v}_i) \in \mathcal{U} \lhd^h \mathcal{V}$, and affine combination coefficients $\alpha_1, \ldots, \alpha_k \in \mathbb{R}, \alpha_1 + \cdots + \alpha_k = 1$, such that

$$(\boldsymbol{u}, \boldsymbol{w}) = \sum_{i=1}^{k} \alpha_i (\boldsymbol{u}_i, h(\boldsymbol{u}_i)\boldsymbol{v}_i) = \left(\sum_{i=1}^{k} \alpha_i \boldsymbol{u}_i, \sum_{i=1}^{k} \alpha_i h(\boldsymbol{u}_i) \boldsymbol{v}_i \right).$$

We break the analysis into two cases.

- If $\sum_{i=1}^{k} \alpha_i h(\boldsymbol{u}_i) = 0$, that is, $\alpha_i h(\boldsymbol{u}_i) = 0$ for all $i = 1, \ldots, k$, consider the point $\boldsymbol{u}^* := \sum_{i=1}^{k} \alpha_i \boldsymbol{u}_i \in \operatorname{aff} \mathcal{U}$. Since h is affine, $h(\boldsymbol{u}^*) = \sum_{i=1}^{k} \alpha_i h(\boldsymbol{u}_i) = 0$. Then,

$$(\boldsymbol{u}, \boldsymbol{w}) = \left(\sum_{i=1}^{k} \alpha_i \boldsymbol{u}_i, \boldsymbol{0} \right) = (\boldsymbol{u}^*, h(\boldsymbol{u}^*)\boldsymbol{w}_1) \in (\operatorname{aff} \mathcal{U}) \overset{h}{\lhd} (\operatorname{aff} \mathcal{V}),$$

where the inclusion holds since $\mathcal{V} \subseteq \operatorname{aff} \mathcal{V}$ (and so, in particular, $\boldsymbol{w}_1 \in \operatorname{aff} \mathcal{V}$).

- Otherwise, $\sum_{i=1}^{k} h(\boldsymbol{u}_i) > 0$. Then,

$$(\boldsymbol{u},\boldsymbol{w}) = \left(\sum_{i=1}^{k}\alpha_i\boldsymbol{u}_i, \sum_{i=1}^{k}\alpha_i h(\boldsymbol{u}_i)\boldsymbol{v}_i\right) = \left(\sum_{i=1}^{k}\alpha_i\boldsymbol{u}_i, \left(\sum_{i=1}^{k}\alpha_i h(\boldsymbol{u}_i)\right)\sum_{i=1}^{k}\frac{\alpha_i h(\boldsymbol{u}_i)}{\sum_{i=1}^{k}\alpha_i h(\boldsymbol{u}_i)}\boldsymbol{v}_i\right)$$
$$= \left(\sum_{i=1}^{k}\alpha_i\boldsymbol{u}_i, h\left(\sum_{i=1}^{k}\alpha_i\boldsymbol{u}_i\right)\sum_{i=1}^{k}\frac{\alpha_i h(\boldsymbol{u}_i)}{\sum_{i=1}^{k}\alpha_i h(\boldsymbol{u}_i)}\boldsymbol{v}_i\right) \in (\operatorname{aff}\mathcal{U}) \overset{h}{\lhd} (\operatorname{aff}\mathcal{V}),$$

where the last equality follows from the affinity of h, and the inclusion follows from the fact that the coefficients $\alpha_i h(\boldsymbol{u}_i)/(\sum_{i=1}^{k}\alpha h(\boldsymbol{u}_i))$ sum to 1 (*i.e.*, they are affine combination coefficients).

($\supseteq$) Let $\boldsymbol{u}\in\operatorname{aff}\mathcal{U}$ and $\boldsymbol{v}\in\operatorname{aff}\mathcal{V}$. We argue that $(\boldsymbol{u}, h(\boldsymbol{u})\boldsymbol{v}) \in \operatorname{aff}(\mathcal{U}\lhd^h\mathcal{V})$. By definition of affine hull,

$$\boldsymbol{u} = \sum_{i=1}^{k_u}\alpha_i\boldsymbol{u}_i; \qquad \boldsymbol{v} = \sum_{j=1}^{k_v}\beta_j\boldsymbol{v}_j$$

for appropriate $k_u, k_v \in \mathbb{N}_{>0}$, affine combination coefficients α_i, β_j, and points $\boldsymbol{u}_i\in\mathcal{U}$, $\boldsymbol{v}_j\in\mathcal{V}$. Consider now the $k_u\cdot k_v$ vectors and coefficients

$$\boldsymbol{x}_{ij} := (\boldsymbol{u}_i, h(\boldsymbol{u}_i)\boldsymbol{v}_j) \in \mathcal{U}\overset{h}{\lhd}\mathcal{V}, \quad \gamma_{ij} := \alpha_i\beta_j \qquad \forall i\in\{1,\dots,k_u\},\ j\in\{1,\dots,k_v\}.$$

The coefficients γ_{ij} are valid affine combination coefficients, since $\sum_{ij}\gamma_{ij} = \sum_{i=1}^{k_u}\sum_{j=1}^{k_v}\alpha_i\beta_j = (\sum_{i=1}^{k_u}\alpha_i)(\sum_{j=1}^{k_v}\beta_j) = 1$. Furthermore, the affine combination $\sum_{ij}\gamma_{ij}\boldsymbol{x}_{ij}$ is such that

$$\operatorname{aff}(\mathcal{U}\overset{h}{\lhd}\mathcal{V}) \ni \sum_{ij}\gamma_{ij}\boldsymbol{x}_{ij} = \sum_{i=1}^{k_u}\sum_{j=1}^{k_v}\alpha_i\beta_j(\boldsymbol{u}_i, h(\boldsymbol{u}_i)\boldsymbol{v}_j)$$
$$= \left(\sum_{i=1}^{k_u}\alpha_i\boldsymbol{u}_i\left(\sum_{j=1}^{k_v}\beta_j\right), \left(\sum_{i=1}^{k_u}\alpha_i h(\boldsymbol{u}_i)\right)\left(\sum_{j=1}^{k_v}\beta_j\boldsymbol{v}_j\right)\right)$$
$$= \left(\sum_{i=1}^{k_u}\alpha_i\boldsymbol{u}_i, h\left(\sum_{i=1}^{k_u}\alpha_i\boldsymbol{u}_i\right)\left(\sum_{j=1}^{k_v}\beta_j\boldsymbol{v}_j\right)\right) = (\boldsymbol{u}, h(\boldsymbol{u})\boldsymbol{v}).$$

The proof is now complete. □

The next lemma builds on the previous results, and shows that scaled extension and relative interior commute. We recall that the the relative interior of a generic set $S\subseteq\mathbb{R}^n$ is the subset

$$\operatorname{relint} S := \big\{\boldsymbol{s}\in S \mid \exists\,\epsilon>0 : N_\epsilon(\boldsymbol{s})\cap\operatorname{aff} S\subseteq S\big\}\subseteq S,$$

where $N_\epsilon(\boldsymbol{s})$ denotes the open ball of radius ϵ centered at $\boldsymbol{s}$, that is, $N_\epsilon(\boldsymbol{s}) := \{\boldsymbol{x}\in\mathbb{R}^n : \|\boldsymbol{x}-\boldsymbol{s}\|_2 < \epsilon\}$.

LEMMA 2. *Let $\mathcal{U}\subset\mathbb{R}^m, \mathcal{V}\subset\mathbb{R}^n$ be bounded sets, and $h:\mathcal{U}\to\mathbb{R}$ be an affine function $h:\boldsymbol{u}\mapsto\boldsymbol{a}^\top\boldsymbol{u}+b$, nonnegative on $\mathcal{U}$ and strictly positive on* $\operatorname{relint}\mathcal{U}$. *Then,*

$$\operatorname{relint}(\mathcal{U}\overset{h}{\lhd}\mathcal{V}) = (\operatorname{relint}\mathcal{U})\overset{h}{\lhd}(\operatorname{relint}\mathcal{V}).$$

Proof. Since $\mathcal{U}$ and $\mathcal{V}$ are bounded, there exists constants

$$M_u := \sup_{\boldsymbol{u}\in\mathcal{U}}\|\boldsymbol{u}\|_2 < \infty, \qquad M_v := \sup_{\boldsymbol{v}\in\mathcal{V}}\|\boldsymbol{v}\|_2 < \infty.$$

Furthermore, note that

$$0 < \zeta := \sup_{\boldsymbol{u}\in\mathcal{U}} h(\boldsymbol{u}) \le M_u\|\boldsymbol{a}\|_2 + b < \infty.$$

We prove the statement by proving the two directions of inclusion separately.

($\subseteq$) Let $\boldsymbol{x} := (\boldsymbol{u}, \boldsymbol{w}) \in \operatorname{relint}(\mathcal{U} \triangleleft^h \mathcal{V})$. We argue that $(\boldsymbol{u}, \boldsymbol{w}) \in (\operatorname{relint}\mathcal{U}) \triangleleft^h (\operatorname{relint}\mathcal{V})$. Since $\operatorname{relint} S \subseteq S$, then $\boldsymbol{w} = h(\boldsymbol{u})\boldsymbol{v}$ for some appropriate $\boldsymbol{v} \in \mathcal{V}$. Furthemore, by definition of relative interior, there exists a real number $\epsilon > 0$ such that

$$N_\epsilon(\boldsymbol{x}) \cap \operatorname{aff}(\mathcal{U} \overset{h}{\triangleleft} \mathcal{V}) \subseteq \mathcal{U} \overset{h}{\triangleleft} \mathcal{V}. \tag{34}$$

Let $\epsilon' := \epsilon/\sqrt{1 + \|\boldsymbol{a}\|_2^2 M_v^2} > 0$, consider any $\boldsymbol{u}' \in B_{\epsilon'}(\boldsymbol{u}) \cap \operatorname{aff}\mathcal{U}$, and let $\boldsymbol{x}' := (\boldsymbol{u}', h(\boldsymbol{u}')\boldsymbol{v})$. The point $\boldsymbol{x}'$ satisfies two properties:

- First, since $\boldsymbol{u}' \in \operatorname{aff}\mathcal{U}$ by hypothesis, and $\mathcal{V} \in \mathcal{V} \subseteq \operatorname{aff}\mathcal{V}$, it follows that

  $$\boldsymbol{x}' = (\boldsymbol{u}', h(\boldsymbol{u}')\boldsymbol{v}) \in (\operatorname{aff}\mathcal{U}) \overset{h}{\triangleleft} (\operatorname{aff}\mathcal{V}) = \operatorname{aff}(\mathcal{U} \overset{h}{\triangleleft} \mathcal{V}),$$

  where the last equality follows from Lemma 4.
- Second,

  $$\begin{aligned} \|\boldsymbol{x}' - \boldsymbol{x}\|_2^2 &= \|(\boldsymbol{u}, h(\boldsymbol{u})\boldsymbol{v}) - (\boldsymbol{u}', h(\boldsymbol{u}')\boldsymbol{v})\|_2^2 = \|\boldsymbol{u} - \boldsymbol{u}'\|_2^2 + (h(\boldsymbol{u}) - h(\boldsymbol{u}'))^2 \|\boldsymbol{v}\|_2^2 \\ &\le \|\boldsymbol{u} - \boldsymbol{u}'\|_2^2 + (\boldsymbol{a}^\top(\boldsymbol{u} - \boldsymbol{u}'))^2 M_v^2 \\ &\le (1 + \|\boldsymbol{a}\|_2^2 M_v^2)\|\boldsymbol{u} - \boldsymbol{u}'\|_2^2 \qquad \text{(apply the Cauchy-Schwarz inequality)} \\ &< \epsilon^2, \end{aligned}$$

  that is, $\boldsymbol{x}' \in N_\epsilon(\boldsymbol{x})$.

Combining the two properties above, we have that $\boldsymbol{x}' \in N_\epsilon(\boldsymbol{x}) \cap \operatorname{aff}(\mathcal{U} \triangleleft^h \mathcal{V})$. So, using (34), we obtain that $\boldsymbol{x}' \in (\mathcal{U} \triangleleft^h \mathcal{V})$, which implies $\boldsymbol{u}' \in \mathcal{U}$. Now, since $\boldsymbol{u}' \in \operatorname{aff}\mathcal{U}$ was an arbitrary point such that $\|\boldsymbol{u}' - \boldsymbol{u}\|_2 < \epsilon'$, we obtain that

$$N_{\epsilon'}(\boldsymbol{u}) \cap \operatorname{aff}\mathcal{U} \subseteq \mathcal{U}, \quad \text{which implies that} \quad \boldsymbol{u} \in \operatorname{relint}\mathcal{U}. \tag{35}$$

We prove that $\boldsymbol{v} \in \operatorname{relint}\mathcal{V}$ in a similar fashion. Let $\epsilon'' := \epsilon/\zeta > 0$ and consider any $\boldsymbol{v}'' \in B_{\epsilon''}(\boldsymbol{v}) \cap \operatorname{aff}\mathcal{V}$. The point $\boldsymbol{x}'' := (\boldsymbol{u}, h(\boldsymbol{u})\boldsymbol{v}'')$ belongs to $(\operatorname{aff}\mathcal{U}) \triangleleft^h (\operatorname{aff}\mathcal{V}) = \operatorname{aff}(\mathcal{U} \triangleleft^h \mathcal{V})$, and furthermore

$$\begin{aligned} \|\boldsymbol{x}'' - \boldsymbol{x}\|_2^2 &= \|(\boldsymbol{u}, h(\boldsymbol{u})\boldsymbol{v}) - (\boldsymbol{u}, h(\boldsymbol{u})\boldsymbol{v}'')\|_2^2 = h(\boldsymbol{u})^2 \|\boldsymbol{v} - \boldsymbol{v}'\|_2^2 \le \zeta^2 \|\boldsymbol{v} - \boldsymbol{v}''\|_2^2 \\ &< \epsilon^2. \end{aligned}$$

This shows that

$$\boldsymbol{x}'' \in N_\epsilon(\boldsymbol{x}) \cap \operatorname{aff}(\mathcal{U} \overset{h}{\triangleleft} \mathcal{V}) \subseteq \mathcal{U} \overset{h}{\triangleleft} \mathcal{V}.$$

Since $\boldsymbol{x}'' = (\boldsymbol{u}, h(\boldsymbol{u})\boldsymbol{v}'')$, this implies that $\boldsymbol{v}'' \in \mathcal{V}$. Since $\boldsymbol{v}''$ was arbitrary, we have proved that

$$N_{\epsilon''}(\boldsymbol{v}) \cap \operatorname{aff}\mathcal{V} \subseteq \mathcal{V}, \quad \text{which implies that} \quad \boldsymbol{v} \in \operatorname{relint}\mathcal{V}. \tag{36}$$

Putting (35) and (36) together, we have that

$$\boldsymbol{x} = (\boldsymbol{u}, h(\boldsymbol{u})\boldsymbol{v}) \in (\operatorname{relint}\mathcal{U}) \overset{h}{\triangleleft} (\operatorname{relint}\mathcal{V}),$$

as we wanted to show.

($\supseteq$) Let $\boldsymbol{u} \in \operatorname{relint}\mathcal{U}$ and $\boldsymbol{v} \in \operatorname{relint}\mathcal{V}$. We argue that $(\boldsymbol{u}, h(\boldsymbol{u})\boldsymbol{v}) \in \operatorname{relint}(\mathcal{U} \triangleleft^h \mathcal{V})$. By definition of relevative interior, there exist constants $\epsilon_u, \epsilon_v > 0$ such that

$$N_{\epsilon_u}(\boldsymbol{u}) \cap \operatorname{aff}\mathcal{U} \subseteq \mathcal{U}, \qquad \text{and} \qquad N_{\epsilon_v}(\boldsymbol{v}) \cap \operatorname{aff}\mathcal{V} \subseteq \mathcal{V}. \tag{37}$$

Let

$$\epsilon' := \frac{\min\{\epsilon_u, \epsilon_v\}}{\sqrt{\max\left\{\frac{2}{h(\boldsymbol{u})^2}, 1 + \frac{2\|\boldsymbol{a}\|_2^2 M_v^2}{h(\boldsymbol{u})^2}\right\}}} > 0 \tag{38}$$

and let $\boldsymbol{x}' \in B_{\epsilon'}(\boldsymbol{x}) \cap \operatorname{aff}(\mathcal{U} \triangleleft^h \mathcal{V})$ be arbitrary. Since from Lemma 4 $\operatorname{aff}(\mathcal{U} \triangleleft^h \mathcal{V}) = (\operatorname{aff}\mathcal{U}) \triangleleft^h (\operatorname{aff}\mathcal{V})$, then $\boldsymbol{x}' = (\boldsymbol{u}', h(\boldsymbol{u}')\boldsymbol{v}')$ for appropriate $\boldsymbol{u}' \in \operatorname{aff}\mathcal{U}, \boldsymbol{v}' \in \operatorname{aff}\mathcal{V}$. Now, since $\boldsymbol{u} \in \operatorname{relint}\mathcal{U}$ and h is strictly positive in the relative interior of $\mathcal{U}$ by hypothesis, we can write

$$
\begin{aligned}
\|\boldsymbol{u}-\boldsymbol{u}'\|_2^2 + \|\boldsymbol{v}-\boldsymbol{v}'\|_2^2 &= \|\boldsymbol{u}-\boldsymbol{u}'\|_2^2 + \frac{1}{h(\boldsymbol{u})^2}\|(h(\boldsymbol{u})\boldsymbol{v} - h(\boldsymbol{u}')\boldsymbol{v}') + (h(\boldsymbol{u}') - h(\boldsymbol{u}))\boldsymbol{v}'\|_2^2 \\
&\le \|\boldsymbol{u}-\boldsymbol{u}'\|_2^2 + \frac{2}{h(\boldsymbol{u})^2}\|h(\boldsymbol{u})\boldsymbol{v} - h(\boldsymbol{u}')\boldsymbol{v}'\|_2^2 + \frac{2}{h(\boldsymbol{u})^2}(h(\boldsymbol{u}) - h(\boldsymbol{u}'))^2\|\boldsymbol{v}'\|_2^2 \\
&\le \|\boldsymbol{u}-\boldsymbol{u}'\|_2^2 + \frac{2}{h(\boldsymbol{u})^2}\|h(\boldsymbol{u})\boldsymbol{v} - h(\boldsymbol{u}')\boldsymbol{v}'\|_2^2 + \frac{2\|\boldsymbol{a}\|_2^2 M_v^2}{h(\boldsymbol{u})^2}\|\boldsymbol{u}-\boldsymbol{u}'\|_2^2 \qquad (39) \\
&\le \max\left\{\frac{2}{h(\boldsymbol{u})^2}, 1 + \frac{2\|\boldsymbol{a}\|_2^2 M_v^2}{h(\boldsymbol{u})^2}\right\}\left(\|\boldsymbol{u}-\boldsymbol{u}'\|_2^2 + \|h(\boldsymbol{u})\boldsymbol{v} - h(\boldsymbol{u}')\boldsymbol{v}'\|_2^2\right) \\
&= \max\left\{\frac{2}{h(\boldsymbol{u})^2}, 1 + \frac{2\|\boldsymbol{a}\|_2^2 M_v^2}{h(\boldsymbol{u})^2}\right\}\|(\boldsymbol{u}, h(\boldsymbol{u})\boldsymbol{v}) - (\boldsymbol{u}', h(\boldsymbol{u}')\boldsymbol{v}')\|_2^2 \\
&= \max\left\{\frac{2}{h(\boldsymbol{u})^2}, 1 + \frac{2\|\boldsymbol{a}\|_2^2 M_v^2}{h(\boldsymbol{u})^2}\right\}\|\boldsymbol{x} - \boldsymbol{x}'\|_2^2 < \min\{\epsilon_u, \epsilon_v\}^2, \qquad (40)
\end{aligned}
$$

where we applied the Cauchy-Schwarz inequality in (39), and the last inequality follows from the hypothesis that $\boldsymbol{x}' \in B_{\epsilon'}(\boldsymbol{x})$ together with the definition of ϵ' given in (38). Now, since $\|\boldsymbol{u}-\boldsymbol{u}'\|_2^2$ and $\|\boldsymbol{v}-\boldsymbol{v}'\|_2^2$ are nonnegative quantities, (40) implies that

$$\|\boldsymbol{u}-\boldsymbol{u}'\|_2 < \min\{\epsilon_u, \epsilon_v\}, \qquad \text{and} \qquad \|\boldsymbol{v}-\boldsymbol{v}'\|_2 < \min\{\epsilon_u, \epsilon_v\}.$$

Furthermore, since $\boldsymbol{u}' \in \operatorname{aff}\mathcal{U}, \boldsymbol{v}' \in \operatorname{aff}\mathcal{V}$ by construction, we have that

$$\boldsymbol{u}' \in (B_{\min\{\epsilon_u,\epsilon_v\}}(\boldsymbol{u}) \cap \operatorname{aff}\mathcal{U}) \subseteq (B_{\epsilon_u}(\boldsymbol{u}) \cap \operatorname{aff}\mathcal{U}), \qquad \boldsymbol{v}' \in (B_{\min\{\epsilon_u,\epsilon_v\}}(\boldsymbol{v}) \cap \operatorname{aff}\mathcal{V}) \subseteq (B_{\epsilon_v}(\boldsymbol{v}) \cap \operatorname{aff}\mathcal{V}).$$

Hence, using the hypothesis (37), we obtain that $\boldsymbol{u}' \in \mathcal{U}$ and $\boldsymbol{v}' \in \mathcal{V}$, which implies that

$$\boldsymbol{x}' = (\boldsymbol{u}', h(\boldsymbol{u}')\boldsymbol{v}') \in \mathcal{U} \overset{h}{\triangleleft} \mathcal{V}.$$

Now, since $\boldsymbol{x}' \in B_{\epsilon'}(\boldsymbol{x}) \cap \operatorname{aff}(\mathcal{U} \triangleleft^h \mathcal{V})$ was arbitrary, this implies

$$B_{\epsilon'}(\boldsymbol{x}) \cap \operatorname{aff}(\mathcal{U} \overset{h}{\triangleleft} \mathcal{V}) \subseteq \mathcal{U} \overset{h}{\triangleleft} \mathcal{V},$$

and so $\boldsymbol{x} \in \operatorname{relint}(\mathcal{U} \triangleleft^h \mathcal{V})$, which is what we wanted to show.

□

LEMMA 5. *Let $\mathcal{X}$ satisfy Setup 1, and that $\mathcal{X}_k = \Delta^{s_k}$ is a probability simplex for all k. Then, for any choices of weights $\alpha_k > 0$, the dilated entropy DGF*

$$
\begin{aligned}
\psi : (\boldsymbol{x}_1, \dots, \boldsymbol{x}_n) \mapsto \alpha_1 &\left(\log(s_1) + \sum_{i=1}^{s_1} x_1[i] \log x_1[i]\right) \\
&+ \sum_{k=2}^{n} \alpha_k \left(h_{k-1}(\boldsymbol{x}_1, \dots, \boldsymbol{x}_{k-1}) \log(s_k) + \sum_{i=1}^{s_k} x_k[i] \log \frac{x_k[i]}{h_{k-1}(\boldsymbol{x}_1, \dots, \boldsymbol{x}_{k-1})}\right).
\end{aligned}
$$

coincides on $\mathcal{X}$ with the dilatable global entropy DGF

$$
\begin{aligned}
\tilde{\varphi} : \mathcal{X} \ni (\boldsymbol{x}_1, \dots, \boldsymbol{x}_n) \mapsto \sum_{k=1}^{n} \left(\alpha_k \sum_{i=1}^{s_k} x_k[i] \log x_k[i]\right) - \sum_{k=2}^{n} \alpha_k h_{k-1}(\boldsymbol{x}_1, \dots, \boldsymbol{x}_{k-1}) \log h_{k-1}(\boldsymbol{x}_1, \dots, \boldsymbol{x}_{k-1}) \\
+ \alpha_1 \log s_1 + \sum_{k=2}^{n} \alpha_k h_{k-1}(\boldsymbol{x}_1, \dots, \boldsymbol{x}_{k-1}) \log s_k.
\end{aligned}
$$

Proof. By the properties of the logarithm,

$$x_k[i]\log\left(\frac{x_k[i]}{h_{k-1}(\boldsymbol{x}_1,\dots,\boldsymbol{x}_{k-1})}\right) = x_k[i]\log(x_k[i]) - x_k[i]\log h_{k-1}(\boldsymbol{x}_1,\dots,\boldsymbol{x}_{k-1})$$

for all $k\in\{1,\dots,n\}$ and $i\in\{1,\dots,s_k\}$. Hence, we can write

$$\begin{aligned}\psi(\boldsymbol{x}_1,\dots,\boldsymbol{x}_n) = \alpha_1\left(\log(s_1)+\sum_{i=1}^{s_1}x_1[i]\log x_1[i]\right)+\sum_{k=2}^{n}\alpha_k\, h_{k-1}(\boldsymbol{x}_1,\dots,\boldsymbol{x}_{k-1})\,\log(s_k)\\ +\sum_{k=2}^{n}\left(\alpha_k\sum_{i=1}^{s_k}x_k[i]\log x_k[i]\right)-\sum_{k=1}^{n}\left(\alpha_k\sum_{i=1}^{s_k}x_k[i]\log h_{k-1}(\boldsymbol{x}_1,\dots,\boldsymbol{x}_{k-1})\right).\end{aligned}\tag{41}$$

By noting that the inner summation variable i in the last term only appears in $x_k[i]$, we can write

$$\sum_{k=1}^{n}\left(\alpha_k\sum_{i=1}^{s_k}x_k[i]\log h_{k-1}(\boldsymbol{x}_1,\dots,\boldsymbol{x}_{k-1})\right)=\sum_{k=1}^{n}\left(\alpha_k\left(\sum_{i=1}^{s_k}x_k[i]\right)\log h_{k-1}(\boldsymbol{x}_1,\dots,\boldsymbol{x}_{k-1})\right).$$

By using the fact that the k-th simplex $\mathcal{X}_k$ is scaled by $h_{k-1}(\boldsymbol{x}_1,\dots,\boldsymbol{x}_{k-1})$, we can write

$$\sum_{i=1}^{s_k}x_k[i]=h_{k-1}(\boldsymbol{x}_1,\dots,\boldsymbol{x}_{k-1}),$$

which leads to

$$\sum_{k=1}^{n}\left(\alpha_k\sum_{i=1}^{s_k}x_k[i]\log h_{k-1}(\boldsymbol{x}_1,\dots,\boldsymbol{x}_{k-1})\right)=\sum_{k=1}^{n}\alpha_k\, h_{k-1}(\boldsymbol{x}_1,\dots,\boldsymbol{x}_{k-1})\log h_{k-1}(\boldsymbol{x}_1,\dots,\boldsymbol{x}_{k-1}).$$

Finally, plugging the last equality into (41) yields the statement. ☐

Appendix C: Further Details

LEMMA 3. *The function d_z as defined in Proposition 1 satisfies*

$$\begin{aligned}(\nabla d_z(\boldsymbol{u},\boldsymbol{w})-\nabla d_z(\boldsymbol{u}',\boldsymbol{w}'))^\top\begin{pmatrix}\boldsymbol{u}-\boldsymbol{u}'\\ \boldsymbol{w}-\boldsymbol{w}'\end{pmatrix}\ge\left(\nabla d_u(\boldsymbol{u})-\nabla d_u(\boldsymbol{u}')\right)^\top(\boldsymbol{u}-\boldsymbol{u}')\\ +\alpha_v h\left(\frac{\boldsymbol{u}+\boldsymbol{u}'}{2}\right)\left\|\frac{\boldsymbol{w}}{h(\boldsymbol{u})}-\frac{\boldsymbol{w}'}{h(\boldsymbol{u})}\right\|_2^2\end{aligned}\tag{18}$$

for all $(\boldsymbol{u},\boldsymbol{w}),(\boldsymbol{u}',\boldsymbol{w}')\in\operatorname{relint}\mathcal{Z}=(\operatorname{relint}\mathcal{U})\lhd^h(\operatorname{relint}\mathcal{V})$.

Proof. Using the expression (17) for the gradient of d_z, for all $(\boldsymbol{u},\boldsymbol{v}),(\boldsymbol{u}',\boldsymbol{v}')\in\operatorname{relint}\mathcal{Z}$ we have

$$\begin{aligned}\nabla d_z(\boldsymbol{u},\boldsymbol{v})^\top\begin{pmatrix}\boldsymbol{u}'\\ \boldsymbol{v}'\end{pmatrix}&=\left(\nabla d_u(\boldsymbol{u})+\alpha_v\left(d_v\left(\frac{\boldsymbol{v}}{h(\boldsymbol{u})}\right)-\nabla d_v\left(\frac{\boldsymbol{v}}{h(\boldsymbol{u})}\right)^\top\frac{\boldsymbol{v}}{h(\boldsymbol{u})}\right)\boldsymbol{a}\right)^\top\boldsymbol{u}'+\left(\alpha_v\nabla d_v\left(\frac{\boldsymbol{v}}{h(\boldsymbol{u})}\right)\right)^\top\boldsymbol{v}'\\ &=\nabla d_u(\boldsymbol{u})^\top\boldsymbol{u}'+\alpha_v d_v\left(\frac{\boldsymbol{v}}{h(\boldsymbol{u})}\right)h(\boldsymbol{u}')-\alpha_v\nabla d_v\left(\frac{\boldsymbol{v}}{h(\boldsymbol{u})}\right)^\top\frac{\boldsymbol{v}}{h(\boldsymbol{u})}h(\boldsymbol{u}')+\alpha_v\nabla d_v\left(\frac{\boldsymbol{v}}{h(\boldsymbol{u})}\right)^\top\boldsymbol{v}'\\ &=\nabla d_u(\boldsymbol{u})^\top\boldsymbol{u}'-\alpha_v h(\boldsymbol{u}')\left[-d_v\left(\frac{\boldsymbol{v}}{h(\boldsymbol{u})}\right)-\nabla d_v\left(\frac{\boldsymbol{v}}{h(\boldsymbol{u})}\right)^\top\left(\frac{\boldsymbol{v}'}{h(\boldsymbol{u}')}-\frac{\boldsymbol{v}}{h(\boldsymbol{u})}\right)\right],\end{aligned}\tag{42}$$

where the second equality uses the fact that $h(\boldsymbol{u}') = \boldsymbol{a}^\top \boldsymbol{u}'$ by hypothesis. In particular, when $(\boldsymbol{u}', \boldsymbol{v}') = (\boldsymbol{u}, \boldsymbol{v})$ we have

$$\nabla d_z(\boldsymbol{u}, \boldsymbol{v})^\top \begin{pmatrix} \boldsymbol{u} \\ \boldsymbol{v} \end{pmatrix} = \nabla d_u(\boldsymbol{u})^\top \boldsymbol{u} + \alpha_v h(\boldsymbol{u}) d_v\left(\frac{\boldsymbol{v}}{h(\boldsymbol{u})}\right), \tag{43}$$

and therefore

$$\begin{aligned}
\nabla d_z(\boldsymbol{u}, \boldsymbol{v})^\top \begin{pmatrix} \boldsymbol{u} - \boldsymbol{u}' \\ \boldsymbol{v} - \boldsymbol{v}' \end{pmatrix} &= \nabla d_u(\boldsymbol{u})^\top (\boldsymbol{u} - \boldsymbol{u}') + \alpha_v h(\boldsymbol{u}) d_v\left(\frac{\boldsymbol{v}}{h(\boldsymbol{u})}\right) \\
&\qquad + \alpha_v h(\boldsymbol{u}') \left[-d_v\left(\frac{\boldsymbol{v}}{h(\boldsymbol{u})}\right) - \nabla d_v\left(\frac{\boldsymbol{v}}{h(\boldsymbol{u})}\right)^\top \left(\frac{\boldsymbol{v}'}{h(\boldsymbol{u}')} - \frac{\boldsymbol{v}}{h(\boldsymbol{u})}\right) \right] \\
&= \nabla d_u(\boldsymbol{u})^\top (\boldsymbol{u} - \boldsymbol{u}') + \alpha_v h(\boldsymbol{u}) d_v\left(\frac{\boldsymbol{v}}{h(\boldsymbol{u})}\right) - \alpha_v h(\boldsymbol{u}') d_v\left(\frac{\boldsymbol{v}'}{h(\boldsymbol{u}')}\right) \\
&\qquad + \alpha_v h(\boldsymbol{u}') \left[d_v\left(\frac{\boldsymbol{v}'}{h(\boldsymbol{u}')}\right) - d_v\left(\frac{\boldsymbol{v}}{h(\boldsymbol{u})}\right) - \nabla d_v\left(\frac{\boldsymbol{v}}{h(\boldsymbol{u})}\right)^\top \left(\frac{\boldsymbol{v}'}{h(\boldsymbol{u}')} - \frac{\boldsymbol{v}}{h(\boldsymbol{u})}\right) \right].
\end{aligned}$$

Since $\alpha_v \geq 0$ and h is a nonnegative function by hypothesis, using the strong convexity hypothesis on d_v yields

$$\begin{aligned}
\nabla d_z(\boldsymbol{u}, \boldsymbol{v})^\top \begin{pmatrix} \boldsymbol{u} - \boldsymbol{u}' \\ \boldsymbol{v} - \boldsymbol{v}' \end{pmatrix} &\geq \nabla d_u(\boldsymbol{u})^\top (\boldsymbol{u} - \boldsymbol{u}') + \alpha_v h(\boldsymbol{u}) d_v\left(\frac{\boldsymbol{v}}{h(\boldsymbol{u})}\right) - \alpha_v h(\boldsymbol{u}') d_v\left(\frac{\boldsymbol{v}'}{h(\boldsymbol{u}')}\right) \\
&\qquad + \alpha_v \frac{h(\boldsymbol{u}')}{2} \left\| \frac{\boldsymbol{v}'}{h(\boldsymbol{u}')} - \frac{\boldsymbol{v}}{h(\boldsymbol{u})} \right\|^2.
\end{aligned} \tag{44}$$

Symmetrically,

$$\begin{aligned}
\nabla d_z(\boldsymbol{u}', \boldsymbol{v}')^\top \begin{pmatrix} \boldsymbol{u}' - \boldsymbol{u} \\ \boldsymbol{v}' - \boldsymbol{v} \end{pmatrix} &\geq \nabla d_u(\boldsymbol{u}')^\top (\boldsymbol{u}' - \boldsymbol{u}) + \alpha_v h(\boldsymbol{u}') d_v\left(\frac{\boldsymbol{v}'}{h(\boldsymbol{u}')}\right) - \alpha_v h(\boldsymbol{u}) d_v\left(\frac{\boldsymbol{v}}{h(\boldsymbol{u})}\right) \\
&\qquad + \alpha_v \frac{h(\boldsymbol{u})}{2} \left\| \frac{\boldsymbol{v}'}{h(\boldsymbol{u}')} - \frac{\boldsymbol{v}}{h(\boldsymbol{u})} \right\|^2.
\end{aligned} \tag{45}$$

Adding (45) and (44) together we obtain

$$\begin{aligned}
(\nabla d_z(\boldsymbol{u}, \boldsymbol{v}) - \nabla d_z(\boldsymbol{u}', \boldsymbol{v}'))^\top \begin{pmatrix} \boldsymbol{u} - \boldsymbol{u}' \\ \boldsymbol{v} - \boldsymbol{v}' \end{pmatrix} &\geq \big(\nabla d_u(\boldsymbol{u}) - \nabla d_u(\boldsymbol{u}')\big)^\top (\boldsymbol{u} - \boldsymbol{u}') \\
&\qquad + \alpha_v \left(\frac{h(\boldsymbol{u})}{2} + \frac{h(\boldsymbol{u}')}{2} \right) \left\| \frac{\boldsymbol{v}}{h(\boldsymbol{u})} - \frac{\boldsymbol{v}'}{h(\boldsymbol{u})} \right\|^2 \\
&= \big(\nabla d_u(\boldsymbol{u}) - \nabla d_u(\boldsymbol{u}')\big)^\top (\boldsymbol{u} - \boldsymbol{u}') \\
&\qquad + \alpha_v h\left(\frac{\boldsymbol{u} + \boldsymbol{u}'}{2} \right) \left\| \frac{\boldsymbol{v}}{h(\boldsymbol{u})} - \frac{\boldsymbol{v}'}{h(\boldsymbol{u})} \right\|^2,
\end{aligned}$$

which is what we wanted to prove. □

PROPOSITION 3. *Let $\mathcal{X}$ be obtained via scaled extension of simplex domains $\mathcal{X}_k = \Delta^{s_k}$. Then, the dilated entropy DGF (Definition 7) for $\mathcal{X}$ is 1-strongly convex with respect to the Euclidean norm, and $(1/M_{\mathcal{X}})$-strongly convex with respect to the ℓ_1 norm.*

Proof. Strong convexity with modulus 1 with respect to the Euclidean norm follows directly from Corollary 1. So, we only need to establish $(1/M_{\mathcal{X}})$-strong convexity with respect to the ℓ_1 norm.

Let ψ be the dilated entropy DGF for $\mathcal{X}$, as defined in Definition 7. As a first step, we will show by induction on n that for all $\boldsymbol{m} = (\boldsymbol{m}_1, \ldots, \boldsymbol{m}_n) \in \mathbb{R}^{s_1} \times \cdots \times \mathbb{R}^{s_n}$ and $\boldsymbol{x} = (\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n) \in \operatorname{relint} \mathcal{X} = (\operatorname{relint} \mathcal{X}_1) \lhd^{h_1} \ldots \lhd^{h_{n-1}} \mathcal{X}_n$,

$$\boldsymbol{m}^\top \nabla^2 \psi(\boldsymbol{x}) \boldsymbol{m} \ge \sum_{k=1}^{n} \sum_{i=1}^{s_k} \left(\frac{\alpha_k}{2} - \sum_{p=k}^{n-1} \alpha_{p+1} \|\boldsymbol{a}_p\|_0 \, a_{p,k}[i] \right) \frac{m_k[i]^2}{x_k[i]}. \tag{46}$$

Base case The base case corresponds to the case where $n = 1$ an $\mathcal{X} = \Delta^{s_1}$, that is, no scaled extension is performed. In that case,

$$d : \mathcal{X} \ni \boldsymbol{x}_1 \to \alpha_1 d_1(\boldsymbol{x}_1) = \alpha_1 \log(s_1) + \alpha_1 \sum_{i=1}^{s_1} x_1[i] \log x_1[i].$$

Since $\nabla^2 d(\boldsymbol{x}_1) = \operatorname{diag}\left(\frac{1}{x_1[i]} : i = 1, \ldots, s_1\right)$, we have that for all $\boldsymbol{m} \in \mathbb{R}^{s_1}$ and all $\boldsymbol{x} \in \operatorname{relint} \mathcal{X}$,

$$\boldsymbol{m}^\top \nabla^2 d(\boldsymbol{x}) \boldsymbol{m} = \sum_{i=1}^{s_1} \alpha_1 \frac{m[i]^2}{x[i]} \ge \sum_{i=1}^{s_1} \frac{\alpha_1}{2} \frac{m[i]^2}{x[i]},$$

which satisfies the inductive statement.

Inductive step Let $\mathcal{U} := \mathcal{X}_1 \lhd^{h_1} \ldots \lhd^{h_{n-2}} \mathcal{X}_{n-1}$, and let $\tilde{d}$ be the dilated entropy DGF constructed for $\mathcal{U}$. We will show that the inductive statement continues to hold after one further application of the inductive DGF construction is performed, that is, for the dilated entropy DGF

$$d : (\mathcal{U} \overset{h_{n-1}}{\lhd} \Delta^{s_n}) \ni (\boldsymbol{u}, \boldsymbol{w}) \mapsto \tilde{d}(\boldsymbol{u}) + \alpha_n \left(\log(s_n) + \sum_{i=1}^{s_n} w[i] \log \frac{w[i]}{h_{n-1}(\boldsymbol{u})} \right).$$

Note that for all $(\boldsymbol{u}, \boldsymbol{w}) \in (\operatorname{relint} \mathcal{U}) \lhd^{h_{n-1}} (\operatorname{relint} \Delta^{s_n})$,

$$\nabla^2_{\boldsymbol{u}} d(\boldsymbol{u}, \boldsymbol{w}) = \nabla^2 \tilde{d}(\boldsymbol{u}) + \alpha_n \frac{\boldsymbol{a}_{n-1} \boldsymbol{a}_{n-1}^\top}{h_{n-1}(\boldsymbol{u})}, \qquad \nabla^2_{\boldsymbol{w}} d(\boldsymbol{u}, \boldsymbol{w}) = \alpha_n \operatorname{diag}\left(\left\{ \frac{1}{w[i]} : i = 1, \ldots, s_n \right\} \right),$$

and

$$\frac{\partial^2}{\partial u[i] \partial w[j]} d(\boldsymbol{u}, \boldsymbol{w}) = \frac{\partial^2}{\partial w[j] \partial u[i]} d(\boldsymbol{u}, \boldsymbol{w}) = -\alpha_n \frac{a_{n-1}[i]}{h_{n-1}(\boldsymbol{u})} \qquad \begin{array}{l} \forall\ i = 1, \ldots, s_1 + \cdots + s_{n-1}, \\ \quad j = 1, \ldots, s_n. \end{array}$$

Let $\boldsymbol{m} = (\boldsymbol{m}_1, \ldots, \boldsymbol{m}_n) \in \mathbb{R}^{s_1} \times \cdots \times \mathbb{R}^{s_n}$ and $\boldsymbol{x} = (\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n) \in \operatorname{relint} \mathcal{X} = (\operatorname{relint} \mathcal{U}) \lhd^{h_{n-1}} (\operatorname{relint} \Delta^{s_n})$ be arbitrary, and introduce the vector $\tilde{\boldsymbol{x}} := (\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n) \in \operatorname{relint} \mathcal{U}$ and $\tilde{\boldsymbol{m}} := (\boldsymbol{m}_1, \ldots, \boldsymbol{m}_{n-1})$. Then,

$$\begin{aligned} &\boldsymbol{m}^\top \nabla^2 d(\boldsymbol{x}) \boldsymbol{m} \\ &\quad = \tilde{\boldsymbol{m}}^\top \nabla^2 \tilde{d}(\tilde{\boldsymbol{x}}) \tilde{\boldsymbol{m}} + \alpha_n \frac{(\boldsymbol{a}_{n-1}^\top \tilde{\boldsymbol{m}})^2}{h_{n-1}(\tilde{\boldsymbol{x}})} - 2\alpha_n \frac{\boldsymbol{a}_{n-1}^\top \tilde{\boldsymbol{m}}}{h_{n-1}(\tilde{\boldsymbol{x}})} (\boldsymbol{1}^\top \boldsymbol{m}_n) + \alpha_n \sum_{i=1}^{s_n} \frac{m_n[i]^2}{x_n[i]} \\ &\quad = \tilde{\boldsymbol{m}}^\top \nabla^2 \tilde{d}(\tilde{\boldsymbol{x}}) \tilde{\boldsymbol{m}} + \alpha_n \underbrace{\left[\frac{(\boldsymbol{a}_{n-1}^\top \tilde{\boldsymbol{m}})^2}{h_{n-1}(\tilde{\boldsymbol{x}})} - 2 \frac{\boldsymbol{a}_{n-1}^\top \tilde{\boldsymbol{m}}}{h_{n-1}(\tilde{\boldsymbol{x}})} (\boldsymbol{1}^\top \boldsymbol{m}_n) + \frac{1}{2} \sum_{i=1}^{s_n} \frac{m_n[i]^2}{x_n[i]} \right]}_{=:\, \Lambda(\boldsymbol{m}_n)} + \frac{\alpha_n}{2} \sum_{i=1}^{s_n} \frac{m_n[i]^2}{x_n[i]}. \end{aligned} \tag{47}$$

We now lower bound the expression Λ in square brackets by explicitly computing a minimizer for $\boldsymbol{m}_n$. The term is strongly convex in $\boldsymbol{m}_n$, and therefore the minimizer is the only point in $\mathbb{R}^{s_n}$ for which the gradient is the $\mathbf{0}$ vector. Specifically, given the partial derivatives

$$\frac{\partial}{\partial m_n[i]}\left[\frac{(\boldsymbol{a}_{n-1}^\top\tilde{\boldsymbol{m}})^2}{h_{n-1}(\tilde{\boldsymbol{x}})}-2\frac{\boldsymbol{a}_{n-1}^\top\tilde{\boldsymbol{m}}}{h_{n-1}(\tilde{\boldsymbol{x}})}(\mathbf{1}^\top\boldsymbol{m}_n)+\frac{1}{2}\sum_{i=1}^{s_n}\frac{m_n[i]^2}{x_n[i]}\right]=-2\frac{\boldsymbol{a}_{n-1}^\top\tilde{\boldsymbol{m}}}{h_{n-1}(\tilde{\boldsymbol{x}})}+\frac{m_n[i]}{x_n[i]},$$

we find that the minimizer of Λ is the vector $\boldsymbol{m}_n^*$ whose coordinates are

$$m_n^*[i]=2x_n[i]\frac{\boldsymbol{a}_{n-1}^\top\tilde{\boldsymbol{m}}}{h_{n-1}(\tilde{\boldsymbol{x}})}.$$

Evaluating Λ in $\boldsymbol{m}_n^*$ in particular yields

$$\begin{aligned}\Lambda(\boldsymbol{m}_n)&\ge\Lambda(\boldsymbol{m}_n^*)\\&=\frac{(\boldsymbol{a}_{n-1}^\top\tilde{\boldsymbol{m}})^2}{h_{n-1}(\tilde{\boldsymbol{x}})}-2\frac{\boldsymbol{a}_{n-1}^\top\tilde{\boldsymbol{m}}}{h_{n-1}(\tilde{\boldsymbol{x}})}\left(\sum_{i=1}^{s_n}2x_n[i]\frac{\boldsymbol{a}_{n-1}^\top\tilde{\boldsymbol{m}}}{h_{n-1}(\tilde{\boldsymbol{x}})}\right)+\frac{1}{2}\sum_{i=1}^{s_n}\frac{1}{x_n[i]}\left(2x_n[i]\frac{\boldsymbol{a}_{n-1}^\top\tilde{\boldsymbol{m}}}{h_{n-1}(\tilde{\boldsymbol{x}})}\right)^2\\&=\frac{(\boldsymbol{a}_{n-1}^\top\tilde{\boldsymbol{m}})^2}{h_{n-1}(\tilde{\boldsymbol{x}})}-4\frac{(\boldsymbol{a}_{n-1}^\top\tilde{\boldsymbol{m}})^2}{h_{n-1}^2(\tilde{\boldsymbol{x}})}\left(\sum_{i=1}^{s_n}x_n[i]\right)+2\frac{(\boldsymbol{a}_{n-1}^\top\tilde{\boldsymbol{m}})^2}{h_{n-1}^2(\tilde{\boldsymbol{x}})}\left(\sum_{i=1}^{s_n}x_n[i]\right).\end{aligned}$$

Finally, note that by definition of scaled extension, $\sum_{i=1}^{s_n}x_n[i]=h_{n-1}(\tilde{\boldsymbol{x}})$. So,

$$\Lambda(\boldsymbol{m}_n)\ge\Lambda(\boldsymbol{m}_n^*)=-\frac{(\boldsymbol{a}_{n-1}^\top\tilde{\boldsymbol{m}})^2}{h_{n-1}(\tilde{\boldsymbol{x}})}.\tag{48}$$

Plugging (48) into (47) yields

$$\begin{aligned}\boldsymbol{m}^\top\nabla^2 d(\boldsymbol{x})\boldsymbol{m}&\ge\tilde{\boldsymbol{m}}^\top\nabla^2\tilde{d}(\tilde{\boldsymbol{x}})\tilde{\boldsymbol{m}}-\alpha_n\frac{(\boldsymbol{a}_{n-1}^\top\tilde{\boldsymbol{m}})^2}{h_{n-1}(\tilde{\boldsymbol{x}})}+\frac{\alpha_n}{2}\sum_{i=1}^{s_n}\frac{m_n[i]^2}{x_n[i]}\\&=\tilde{\boldsymbol{m}}^\top\nabla^2\tilde{d}(\tilde{\boldsymbol{x}})\tilde{\boldsymbol{m}}-\alpha_n\frac{(\boldsymbol{a}_{n-1}^\top\tilde{\boldsymbol{m}})^2}{\boldsymbol{a}_{n-1}^\top\tilde{\boldsymbol{x}}}+\frac{\alpha_n}{2}\sum_{i=1}^{s_n}\frac{m_n[i]^2}{x_n[i]},\end{aligned}\tag{49}$$

where the equality follows from expanding the definition of h_{n-1}. At this point we upper bound the fraction $\frac{(\boldsymbol{a}_{n-1}^\top\tilde{\boldsymbol{m}})^2}{\boldsymbol{a}_{n-1}^\top\tilde{\boldsymbol{x}}}$. First, using the Cauchy-Schwarz inequality,

$$(\boldsymbol{a}_{n-1}^\top\tilde{\boldsymbol{m}})^2\le\|\boldsymbol{a}_{n-1}\|_0\sum_{q=1}^{n-1}\sum_{i=1}^{s_q}a_{n-1,q}^2[i]\,m_q[i]^2.$$

So,

$$\begin{aligned}\frac{(\boldsymbol{a}_{n-1}^\top\tilde{\boldsymbol{m}})^2}{\boldsymbol{a}_{n-1}^\top\tilde{\boldsymbol{x}}}&=\frac{(\boldsymbol{a}_{n-1}^\top\tilde{\boldsymbol{m}})^2}{\sum_{q=1}^{n-1}\sum_{i=1}^{s_q}a_{n-1,q}[i]\,x_q[i]}\\&\le\|\boldsymbol{a}_{n-1}\|_0\frac{\sum_{q=1}^{n-1}\sum_{i=1}^{s_q}a_{n-1,q}^2[i]\,m_q[i]^2}{\sum_{q=1}^{n-1}\sum_{i=1}^{s_q}a_{n-1,q}[i]\,x_q[i]}.\end{aligned}$$

Since the denominator is positive by the assumption that h_{n-1} is a nonnegative function, we can further upper bound a fraction of sums of the form $(\sum c_i)/(\sum d_i)$ with the sum of fractions $\sum(c_i/d_i)$, and obtain

$$\frac{(\boldsymbol{a}_{n-1}^\top\tilde{\boldsymbol{m}})^2}{\boldsymbol{a}_{n-1}^\top\tilde{\boldsymbol{x}}}\le\|\boldsymbol{a}_{n-1}\|_0\sum_{q=1}^{n-1}\sum_{i=1}^{s_q}\frac{a_{n-1,q}[i]\,m_q[i]^2}{x_q[i]}.\tag{50}$$

Finally, plugging (50) into (49), we obtain

$$\boldsymbol{m}^\top \nabla^2 d(\boldsymbol{x})\boldsymbol{m} \ge \tilde{\boldsymbol{m}}^\top \nabla^2 \tilde{d}(\tilde{\boldsymbol{x}})\tilde{\boldsymbol{m}} - \left(\alpha_n \|\boldsymbol{a}_{n-1}\|_0 \sum_{q=1}^{n-1}\sum_{i=1}^{s_q} \frac{a_{n-1,q}[i]\, m_q[i]^2}{x_q[i]}\right) + \frac{\alpha_n}{2}\sum_{i=1}^{s_n}\frac{m_n[i]^2}{x_n[i]}.$$

Substituting the inductive hypothesis, we find

$$\tilde{\boldsymbol{m}}^\top \nabla^2 \tilde{d}(\tilde{\boldsymbol{x}})\tilde{\boldsymbol{m}} \ge \sum_{k=1}^{n-1}\sum_{i=1}^{s_k}\left(\frac{\alpha_k}{2} - \sum_{p=k}^{n-2}\alpha_{p+1}\|\boldsymbol{a}_p\|_0\, a_{p,k}[i]\right)\frac{m_k[i]^2}{x_k[i]}$$

and consolidating terms completes the inductive proof.

Plugging in the definition of the weights α_k defined in Corollary 1 into (46) yields

$$\boldsymbol{m}^\top \nabla^2 \psi(\boldsymbol{x})\boldsymbol{m} \ge \sum_{k=1}^{n}\sum_{i=1}\frac{m_k[i]^2}{x_k[i]} \qquad \forall \boldsymbol{m}\in\mathbb{R}^{s_1+\cdots+s_n}, \boldsymbol{x}\in \operatorname{relint}\mathcal{X}.$$

Hence, using the Cauchy-Schwarz inequality,

$$\begin{aligned}\|\boldsymbol{m}\|_1^2 = \left(\sum_{k=1}^{n}\sum_{i=1}^{s_k} m_k[i]\right)^2 &= \left(\sum_{k=1}^{n}\sum_{i=1}^{s_k}\frac{m_k[i]}{\sqrt{x_k[i]}}\sqrt{x_k[i]}\right)^2 \\ &\le \left(\sum_{k=1}^{n}\sum_{i=1}^{s_k}\frac{m_k[i]^2}{x_k[i]}\right)\left(\sum_{k=1}^{n}\sum_{i=1}^{s_k}x_k[i]\right) \le M_{\mathcal{X}}\boldsymbol{m}^\top\nabla^2\psi(\boldsymbol{x})\boldsymbol{m},\end{aligned}$$

which shows that ψ is $(1/M_{\mathcal{X}})$-strongly convex on relint $\mathcal{X}$ with respect to the ℓ_1 norm. ☐

Appendix D: Detailed Description of Game Instances Used in Numerical Experiments

Here we describe each of the games that we consider in the experimental section of the paper.

- ***Kuhn poker*** is a standard benchmark in the EFG-solving community (Kuhn 1950). In Kuhn poker, each player puts an ante worth 1 into the pot. Each player is then privately dealt one card from a deck that contains 3 unique cards (Jack, Queen, King). Then, a single round of betting then occurs, with the following dynamics. First, Player 1 decides to either check or bet 1. Then,
  - If Player 1 checks Player 2 can check or raise 1.
    - If Player 2 checks a showdown occurs; if Player 2 raises Player 1 can fold or call.
      - If Player 1 folds Player 2 takes the pot; if Player 1 calls a showdown occurs.
  - If Player 1 raises Player 2 can fold or call.
    - If Player 2 folds Player 1 takes the pot; if Player 2 calls a showdown occurs.

  When a showdown occurs, the player with the higher card wins the pot and the game immediately ends.
- ***Leduc poker*** is another standard benchmark in the EFG-solving community (Southey et al. 2005). The game is played with a deck of R unique cards, each of which appears exactly twice in the deck. The game is composed of two rounds. In the first round, each player places an ante of 1 in the pot and is dealt a single private card. A round of betting then takes place, with Player 1 acting first. At most two bets are allowed per player. Then, a card is is revealed face up and another round of betting takes place, with the same dynamics described above. After the two betting round, if one of the players has a pair with the public card, that player wins the pot. Otherwise, the player with the higher card wins the pot. All bets in the first round are worth 1, while all bets in the second round are 2.

- ***Goofspiel*** is another popular benchmark game, originally proposed by Ross (1971). It is a two-player card game, employing three identical decks of k cards each whose values range from 1 to k. At the beginning of the game, each player gets dealt a full deck as their hand, and the third deck (the "prize" deck) is shuffled and put face down on the board. In each turn, the topmost card from the prize deck is revealed. Then, each player privately picks a card from their hand. This card acts as a bid to win the card that was just revealed from the prize deck. The selected cards are simultaneously revealed, and the highest one wins the prize card.

  In the zero-sum version of the game, if the players' played cards are equal, the prize card is split. In the general-sum version of the game, denoted "General-sum Goofspiel" and used in the experiments on NFCCE, the prize card is thrown out on tie. Either way, the players' scores are computed as the sum of the values of the prize cards they have won.
- ***Pursuit-evasion*** is a security-inspired pursuit-evasion game played on the graph shown in Figure 6. It is a zero-sum variant of the one used by Kroer et al. (2018a), and a similar search game has been considered by Bošanskỳ et al. (2014) and Bošanskỳ and Čermák (2015).

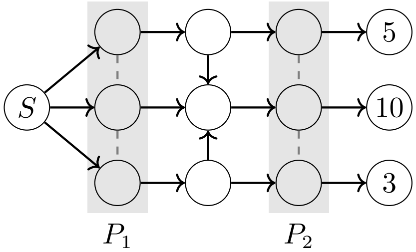

Figure 6 The graph on which the search game is played.

  In each turn, the attacker and the defender act simultaneously. The defender controls two patrols, one per each respective patrol areas labeled P_1 and P_2. Each patrol can move by one step along the grey dashed lines, or stay in place. The attacker starts from the leftmost node (labeled S) and at each turn can move to any node adjacent to its current position by following the black directed edges. The attacker can also choose to wait in place for a time step in order to hide all their traces. If a patrol visits a node that was previously visited by the attacker, and the attacker did not wait to clean up their traces, they will see that the attacker was there. The goal of the attacker is to reach any of the rightmost nodes, whose corresponding payoffs are 5, 10, or 3, respectively, as indicated in Figure 6. If at any time the attacker and any patrol meet at the same node, the attacker is loses the game, which leads to a payoff of -1 for the attacker and of 1 for the defender. The game times out after m simultaneous moves, in which case both players defender receive payoffs 0.
- ***Battleship*** is a parametric version of a classic board game, where two competing fleets take turns shooting at each other (Farina et al. 2019c). At the beginning of the game, the players take turns at secretly placing a set of ships on separate grids (one for each player) of size 3×2. Each ship has size 2 (measured in terms of contiguous grid cells) and a value of 4, and must be placed so that all the cells that make up the ship are fully contained within each player's grids and do not overlap with any other ship that the player has already positioned on the grid. After all ships have been placed. the players take turns at firing at their opponent. Ships that have been hit at all their cells are considered sunk. The game continues until either one player has sunk all of the opponent's ships, or each player has completed R shots. At the end of the game, each player's payoff is calculated as the sum of the values of the opponent's ships that were sunk, minus the sum of the values of ships which that player has lost.

  In the general-sum variant we consider in the NFCCE experiments, we set $R = 3$, and furthermore we set each player's payoff is calculated as the sum of the values of the opponent's ships that were sunk, minus the sum of the values of ships which that player has lost *times two*. This modification makes the game general-sum, and makes the players more risk-averse. Because of that, it was observed by Farina et al. (2019c) that the introduction of a mediator in the game (through the correlated solution concept) enables to players to reach equilibrium states with significantly larger social welfare.

- ***Liar's dice*** is another standard benchmark in the EFG-solving community (Lisý et al. 2015). In our instantiation, each of the two players initially privately rolls an unbiased 6-face die. The first player begins bidding, announcing any face value up to 6 and the minimum number of dice that the player believes are showing that value among the dice of both players. Then, each player has two choices during their turn: to make a higher bid, or to challenge the previous bid by declaring the previous bidder a "liar". A bid is higher than the previous one if either the face value is higher, or the number of dice is higher. If the current player challenges the previous bid, all dice are revealed. If the bid is valid, the last bidder wins and obtains a reward of $+1$ while the challenger obtains a negative payoff of -1. Otherwise, the challenger wins and gets reward $+1$, and the last bidder obtains reward of -1.
- ***Sheriff*** The Sheriff game is inspired by the Sheriff of Nottingham board game and was introduced by Farina et al. (2019c) as a benchmark game for correlated solution concepts in extensive-form game. Player 1 (the ”smuggler”) selects the number of illegal items to be placed in the cargo (in our case, between 0 and 3). The selected number is unknown to Player 2 (the ”sheriff”).

  Then, the game proceeds for 3 bargaining rounds. In each round, the following happens:
  - The smuggler selects an integer bribe amount, in the range 0 to 3 (inclusive). The selected amount is public information. However, the smuggler does not actually give money to the sheriff, unless this is the final round.
  - Then, the sheriff tells the smuggler whether he is planning to inspect the cargo. However, no cargo is actually inspected other than in the final round. The sheriff can change his mind in later rounds, except for the final round.