A Coevolutionary process

Algorithm 4: Coevolutionary Process

Require: Population size $\lambda \in \mathbb{N}$ and search spaces $\mathcal{X}$ and $\mathcal{Y}$.
Require: Initial populations $P_0 \in \mathcal{X}^\lambda$ and $Q_0 \in \mathcal{Y}^\lambda$.
for *each generation number* $t \in \mathbb{N}_0$ **do**
  for *each interaction number* $i \in [\lambda]$ **do**
    Sample an interaction $(x, y) \sim \mathcal{D}(P_t, Q_t)$;
    Set $P_{t+1}(i) := x$ and $Q_{t+1}(i) := y$;

B Helper Lemmas

Theorem 3 ([23]). *Let $X_1, \ldots, X_n$ be independent Poisson trials such that, for $1 \le i \le n$, $\Pr[X_i = 1] = p_i$, where $0 < p_i < 1$. Then, for $X := \sum_{i=1}^n X_i$, and any $\delta \in (0, 1)$,*

$$\Pr[X < (1-\delta)\mathbb{E}[X]] < \exp\left(-\frac{\delta^2 \mathbb{E}[X]}{2}\right).$$

Lemma 5. *Let $a > 1$, $b > 1$, $0 < x \le 1$, $0 < y \le 1$, $0 < c < 1$, $xy \le c$. Let $f(x, y) = (a - (a-1)x)(b - (b-1)y)$. The minimum of $f(x, y)$ with respect to x and y is*

1. $(1-c)(b-1)+1$ *if* $a \ge b$
2. $(1-c)(a-1)+1$ *if* $b \ge a$

Proof. Note that $(a-(a-1)x)(b-(b-1)y) = ab(1-(a-1)x/a)(1-(b-1)y/b)$. Therefore, the minimum of $(a-(a-1)x)(b-(b-1)y)$ is attained by the same x and y as $(1-(a-1)x/a)(1-(b-1)y/b)$, since $a, b > 1$. Let $\alpha := (a-1)/a > 0$ and $\beta := (b-1)/b > 0$.

Now we show that the minimum is attained with $xy = c$ by contradiction. Assume that the minimum is attained for some x_0, y_0 with $x_0 y_0 < c$. For any x_1 y_1 with $x_1 y_1 = c$ there exists an $\varepsilon > 0$ such that either $x_1 \ge x_0 + \varepsilon$, $y_1 \ge y_0 + \varepsilon$ or both. For simplicity lets assume that there exist a $x_1 = x_0 + \varepsilon$ and $y_0 = y_1$. Then,

$$(1-\alpha x_1)(1-\beta y_1) = (1-\alpha(x_0+\varepsilon))(1-\beta y_1) < (1-\alpha(x_0))(1-\beta y_0),$$

which contradicts the statement that the minimum is attained for $x_0 y_0 < c$.

Since the minimum is attained by $xy = c$ then $x, y \ge c$ and

$$\arg\min_{x,y}\{(1-\alpha x)(1-\beta y)\} = \arg\min_{x,y}\{1-\alpha x-\beta y+c\alpha\beta\} = \arg\max_{x,y}\{\alpha x+\beta y\}.$$

Now, we find the extrema of $\alpha x + \beta y$. Given that $xy = c$ then $y = c/x$ and

$$\alpha x + \beta y = \alpha x + \frac{\beta c}{x}.$$

Let $g(x) = \alpha x + \frac{\beta c}{x}$, then $g'(x) = \alpha - \frac{\beta c}{x^2}$ and the extrema are found where $g'(x) = 0$. These are attained for $x = \pm\sqrt{\frac{\beta c}{\alpha}}$. $x \geq 0$ by assumption, therefore the only extremum is $x = \sqrt{\frac{\beta c}{\alpha}}$ and since $g''(x) = \frac{2\beta c}{x^3} > 0$ it is a minimum. Hence, the maximum of $\alpha x + \beta y$ is either $x = 1$ and $y = c$ or $x = c$ and $y = 1$.

Substituting these values of x and y in the original function $f(x, y) = (a - (a-1)x)(b - (b-1)y)$ we can see that if $a > b$, $x = 1$ $y = c$ gives the minimum and $x = c$ $y = 1$ gives the minimum otherwise.

C Omitted Lemmas and Proofs

This appendix contains the lemmas and proofs omitted from the main part.

Theorem 4 (Adapted from [18]). *Given subsets $A_j \subseteq \mathcal{X}$, $B_j \subseteq \mathcal{Y}$ for $j \in [m]$, define $T := \min\{t \mid (P_t \times Q_t) \cap (A_m \times B_m) \neq \emptyset\}$, where for all $t \in \mathbb{N}$, $P_t \in \mathcal{X}^\lambda$ and $Q_t \in \mathcal{Y}^\lambda$ are the populations of Algorithm 4 in generation t. If there exist $z_1, \ldots, z_{m-1}, \delta \in (0, 1]$, and $\gamma_0 \in (0, 1)$ such that for initial populations $|(P_0 \times Q_0) \cap (A_1 \times B_1)| \geq \gamma_0 \lambda^2$, and for any populations $P \in \mathcal{X}^\lambda$ and $Q \in \mathcal{Y}^\lambda$ with "current level" $j := \max\{i \in [m] \mid |(P \times Q) \cap (A_i \times B_i)| \geq \gamma_0 \lambda^2\}$*

(G1) *if $j \in [m-1]$ and $(x, y) \sim \mathcal{D}(P, Q)$ $\Pr[x \in A_{j+1}] \Pr[y \in B_{j+1}] \geq z_j$,*
(G2a) *if $j \in [m-2]$ and all $\gamma \in (0, \gamma_0)$ if $|(P \times Q) \cap (A_{j+1} \times B_{j+1})| \geq \gamma \lambda^2$, then for $(x, y) \sim \mathcal{D}(P, Q)$, $\Pr[x \in A_{j+1}] \Pr[y \in B_{j+1}] \geq (1 + \delta)\gamma$,*
(G2b) *if $j \in [m-1]$ and $(x, y) \sim \mathcal{D}(P, Q)$, $\Pr[x \in A_j] \Pr[y \in B_j] \geq (1 + \delta)\gamma_0$,*
(G3) *and the population size $\lambda \in \mathbb{N}$ satisfies for a sufficiently large constant c', where $z_* := \min_{i \in [m-1]} z_i$, $\lambda \geq c' \log(m/z_*)$,*

then for a constant $c'' > 0$ and any constant $r > 0$,

$$\Pr\left[T \geq c''\left(\lambda^2 m + \sum_{i=1}^{m-1} 1/z_i\right)\right] \leq 1/r.$$

If condition (G2a) is met for $j = m-1$, then for $T' := \min\{t \mid (P_t \times Q_t) \cap (A_m \times B_m) \geq \gamma_0 \lambda^2\}$, a constant $c'' > 0$ and any constant $r > 0$,

$$\Pr\left[T' \geq c''\left(\lambda^2 m + \sum_{i=1}^{m-1} 1/z_i\right)\right] \leq 1/r.$$

To prove Theorem 4 we need a slightly adapted proof of the level-based theorem shown in the supplementary material of [18].

The last part of the original proof of the level-based theorem considers multiple phases. Each phase starts from an arbitrary configuration within the search space $\mathcal{X} \times \mathcal{Y}$ and ends after a fixed number of generations with a certain probability of not finding the optimum within the phase. The overall runtime is obtained by computing the expected number of phases (+1).

In our case we assume $A_1 \times B_1 \neq \mathcal{X} \times \mathcal{Y}$ therefore if the first phase does not find the optimum we cannot restart the analysis from an arbitrary configuration. Therefore, in Theorem 4 we state the runtime as a tail-bound rather than an upper bound on the expected runtime.

In addition, the original proof in the supplementary material of [18] assumed that (G2a) is met for $j = m - 1$, since it only bounded the first time $((P_t \times Q_t) \cap (A_m \times B_m)) \neq \emptyset$. If this is not assumed but instead is a condition of the theorem then we can also show the first time $((P_t \times Q_t) \cap (A_m \times B_m)) \geq \gamma_0 \lambda^2$.

Finally, in [18] the level-based theorem assumes that each generation uses only λ evaluations. Theorem 4 drops this assumption and only bounds the number of generations.

Lemma 1. *Let $r_1 \leq \beta n$, $\beta n \leq r_2$, $s_2 \leq \alpha n$, $\alpha n \leq s_1$ with $\beta n - r_1 = r_2 - \beta n$ and $\alpha n - s_2 = s_1 - \alpha n$. Let $\varepsilon > 0$ be a constant. Let x and y be the offspring created in Lines 15 and 20 of Algorithm 1 with $k = \ell = 2$, and $r_{(0)}$ be the probability of not flipping a bit during mutation. If $v_1, v_3, w_1, w_3 \geq 1 - \frac{1}{\sqrt{2}} + \varepsilon$ and there exist constants $\delta, \delta' > 0$ such that $\frac{1+\delta}{1+(1-\delta')(1+\sqrt{2})\varepsilon} \leq r_{(0)}^2 \leq 1$, then, for all $\gamma \in (0, \delta'/2]$ any population with $P \in \mathcal{X}^\lambda$ and $Q \in \mathcal{Y}^\lambda$ with $|(P \times Q) \cap (A(r_1, r_2) \times B(s_2, s_1))| \geq \gamma\lambda^2$ guarantees that*

$$\Pr[x \in A(r_1, r_2)] \Pr[y \in B(s_2, s_1))] \geq (1 + \delta)\gamma.$$

Proof (Proof of Lemma 1). First, we compute the probability that the algorithm selects a parent from P_t in the level $A(r_1, r_2)$ and later deal with $B(s_2, s_1)$. By Lemma 3.2 in [12] the following conditions (probabilities in parenthesis) result in selecting a parent from P_t in the level $A(r_1, r_2)$:

- Both individuals are sampled in R_1 (p^2).
- The two individuals are sampled in R_0 and R_1 ($2p_0p$). Additionally:
  - The two competitors have $\|y\| < \alpha n$ and $\|y\| > \alpha n$ ($2w_1w_3$).
  - The two competitors have $\|y\| = \alpha n$ and $\|y\| > \alpha n$ ($2w_2w_3$).
  - Both competitors have $\|y\| > \alpha n$ (w_3^2).
- The two individuals are sampled in R_1 and R_2 ($2p(1-p-p_0)$). Additionally:
  - The two competitors have $\|y\| < \alpha n$ and $\|y\| = \alpha n$ ($2w_1w_2$).
  - The two competitors have $\|y\| < \alpha n$ and $\|y\| > \alpha n$ ($2w_1w_3$).
  - Both competitors have $\|y\| < \alpha n$ (w_1^2).

We note that we omitted some conditions, but since we are interested in a lower bound for the probability of selecting a parent in level $A(r_1, r_2)$ this is not a problem. Putting everything together we obtain:

$$\begin{aligned}
p_{sel}(A(r_1,r_2)) &\geq p^2 + 2p_0p\left(2w_1w_3 + w_3^2 + 2w_2w_3\right) \\
&\quad + 2p(1-p-p_0)\left(w_1^2 + 2w_1w_2 + 2w_1w_3\right) \\
&= p^2 + 2p_0p\left(w_3(2w_1+2w_2+w_3)\right) \\
&\quad + 2p(1-p-p_0)\left(w_1(w_1+2w_2+2w_3)\right) \\
&= p^2 + 2p_0p\left(w_3(1+w_1+w_2)\right) \\
&\quad + 2p(1-p-p_0)\left(w_1(1+w_2+w_3)\right) \\
&= p\left(p + 2p_0w_3(2-w_3) + 2(1-p-p_0)w_1(2-w_1)\right) \\
&= p\left(p + p_0\left(4w_3 - 2w_3^2\right) + (1-p-p_0)\left(4w_1 - 2w_1^2\right)\right)
\end{aligned}$$

By assumption both $w_1 \geq 1 - \frac{1}{\sqrt{2}} + \varepsilon$ and $w_3 \geq 1 - \frac{1}{\sqrt{2}} + \varepsilon$, therefore

$$\begin{aligned}
&p_{sel}(A(r_1,r_2)) \\
&\geq p\left(p + (p_0 + (1-p-p_0))\left(4\left(1 - \frac{1}{\sqrt{2}} + \varepsilon\right) - 2\left(1 - \frac{1}{\sqrt{2}} + \varepsilon\right)^2\right)\right) \\
&= p\left(p + (1-p)\left(1 + 2\sqrt{2}\varepsilon - 2\varepsilon^2\right)\right)
\end{aligned}$$

We note that $0 < \varepsilon \leq \frac{1}{\sqrt{2}} - \frac{1}{2}$, otherwise $w_1 + w_3 > 1$. Then $(\sqrt{2}-1)\varepsilon \geq 2\varepsilon^2$ for all possible ε and

$$p_{sel}(A(r_1,r_2)) \geq p\left(p + (1-p)\left(1 + \left(1+\sqrt{2}\right)\varepsilon\right)\right)$$

Let $\kappa := 1 + \left(1+\sqrt{2}\right)\varepsilon$ for simplicity.

$$p_{sel}(A(r_1,r_2)) \geq p\left(p + \kappa(1-p)\right) = p\left(\kappa - (\kappa-1)p\right)$$

Using the same arguments for $p_{sel}(B(s_2,s_1))$ we can obtain

$$p_{sel}(B(s_2,s_1)) \geq q\left(\kappa - (\kappa-1)q\right).$$

Now,

$$\begin{aligned}
&\Pr\left[x \in A(r_1,r_2)\right]\Pr\left[y \in B(s_2,s_1))\right] \\
&\geq p_{sel}(A(r_1,r_2))p_{sel}(B(s_2,s_1))r_{(0)}^2 \\
&\geq pq\left(\kappa - (\kappa-1)p\right)\left(\kappa - (\kappa-1)q\right)r_{(0)}^2
\end{aligned}$$

Now assume that $0 < pq \leq \delta'$, we will deal with $pq > \delta'$ later. Under this restriction by Lemma 5 the minimum value of $\left(\kappa - (\kappa-1)p\right)\left(\kappa - (\kappa-1)q\right)$ is $1 + (1-\delta')(\kappa-1)$, hence

$$\begin{aligned}
&\Pr\left[x \in A(r_1,r_2)\right]\Pr\left[y \in B(s_2,s_1))\right] \\
&\geq p_{sel}(A(r_1,r_2))p_{sel}(B(s_2,s_1))r_{(0)}^2 \\
&\geq pq\left(1 + (1-\delta')(\kappa-1)\right)r_{(0)}^2
\end{aligned}$$

By the assumptions on $r_{(0)}^2$ and $pq \geq \gamma$ this is at least $(1+\delta)\gamma$, proving the claim for $0 < pq \leq \delta'$.

We note that $1 + (1-\delta')(\kappa - 1) \geq 1$ because $\kappa > 1$ and $\delta' < 1$ to fulfill the requirement for $r_{(0)}^2 \leq 1$. If $pq > \delta'$ we pessimistically assume that $1 + (1-\delta')(\kappa-1) = 1$. We note that since $pq \geq \gamma$ and $\gamma \in (0, \delta'/2]$ there is an $\delta' - \delta'/2 < \varepsilon' \leq 1 - \gamma$ such that $pq = \gamma + \varepsilon'$. Then,

$$\begin{aligned}
\Pr[x \in A(r_1, r_2)] \Pr[y \in B(s_2, s_1))] &\geq pq r_{(0)}^2 \\
&= \gamma \left(1 + \frac{\varepsilon'}{\gamma}\right) r_{(0)}^2 \\
&\geq \gamma \left(1 + \frac{\delta' - \delta'/2}{\delta'/2}\right) r_{(0)}^2 \\
&= 2\gamma r_{(0)}^2.
\end{aligned}$$

As before, by the assumptions on $r_{(0)}^2$ this is at least $(1+\delta)\gamma$.

C.1 Lemmas and proofs of Section 5

In the following lemmas we consider a different type of levels than in Lemma 1 where the restrictions on the archive are more lenient and the algorithm would have the opportunity to encounter a diverse set of solutions to build a good archive.

The new levels consider the whole search space for the prey ($B(0,n)$) and for the predator all solutions with less than r_2 1-bits ($A(0, r_2)$). The idea is that an algorithm that starts at level $A(0,n)$ and ends at a level $A(0, \beta n - 1)$ would encounter individuals in $\mathcal{X}$ that have less than βn 1-bits and more than βn 1-bits, allowing it to build a good predator archive V. Figure 5 (b) shows these levels.

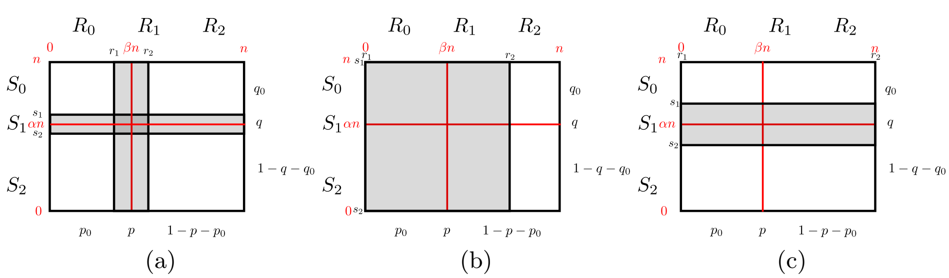

Fig. 5: Levels for Lemma 1 (a), Lemma 6 (b) and Lemma 7 (c) on Bilinear.

Lemma 6 shows that for these levels there are no conditions for the predator archive V and the conditions for the prey archive W and $r_{(0)}$ are easier to meet than in Lemma 1. For example, for $w_1 = w_2 = 1/2$ $r_{(0)} \geq 0.58$ meets the conditions.

Lemma 6. *Let x and y be the offspring created in Lines 15 and 20 of Algorithm 1 with $k = \ell = 2$. If there are constants $\delta, \delta' > 0$ such that any of the following conditions hold:*

1. $r_1 = 0$, $r_2 \geq 0$, $s_2 = 0$, $s_1 = n$, $\frac{1+\delta}{(1-\delta')2\left(w_1^2+w_1w_2+w_2^2/2\right)} \leq r_{(0)} \leq 1$ *and*
   - *(a)* $w_1 + w_2 = 1$,
   - *(b)* $w_1^2 > 1/2$, *or*
   - *(c)* $w_1^2 \leq 1/2$ *and* $w_2 > \sqrt{1 - w_1^2} - w_1$,
2. $r_1 = 0$, $r_2 \geq \beta n$, $s_2 = 0$, $s_1 = n$, $\frac{1+\delta}{(1-\delta')2\left(w_1^2+2w_1w_2+w_2^2/2\right)} \leq r_{(0)} \leq 1$ *and*
   - *(a)* $w_1 + w_2 = 1$,
   - *(b)* $w_1^2 > 1/2$, *or*
   - *(c)* $w_1^2 \leq 1/2$ *and* $w_2 > \sqrt{2w_1^2 + 1} - 2w_1 \geq 1 - \sqrt{2}w_1$

Then, for all $\gamma \in (0, \delta'/2]$ any population with $P \in \mathcal{X}^\lambda$ and $Q \in \mathcal{Y}^\lambda$ with $|(P \times Q) \cap (A(r_1, r_2) \times B(s_2, s_1))| \geq \gamma\lambda^2$

$$\Pr\left[x \in A(r_1, r_2)\right] \Pr\left[y \in B(s_2, s_1))\right] \geq (1 + \delta)\gamma$$

We note that Lemma 6 focuses on levels that cover the whole search space for the population Q and all bit-strings with at most r_2 1-bits for the population P. Equivalent conditions can be proved if we rotate the levels in the search space. That is if we consider all bit-strings with at least r_2 1-bits for the population P, or consider all the search space for the population P and all bit-strings with at most/least s_1 1-bits for the population Q. Due to their similarities (in their conditions and proofs) we omit these.

Proof (Proof of Lemma 6). Since $s_2 = 0$ and $s_1 = n$ then $B(s_2, s_1) = \{0, 1\}^n$ and the probability $\Pr\left[y \in B(s_2, s_1)\right] = 1$. Therefore, we only need to bound $\Pr\left[x \in A(r_1, r_2)\right]$ for both Conditions (1) and (2). We start by proving the statement for Condition (1). By Lemma 3.2 in [12] the following conditions (probabilities in parenthesis) result in selecting a parent from P_t in the level $A(r_1, r_2)$:

- Both individuals are sampled in R_1 (p^2).
- The two individuals are sampled in R_1 and R_2 ($2p(1-p-p_0)$). Additionally:
  - Both competitors have $\|y\| < \alpha n$ (w_1^2).
  - The two competitors have $\|y\| < \alpha n$ and $\|y\| = \alpha n$ and the individual in R_1 is chosen u. a. r. in line 14 ($w_1 w_2$).
  - Both competitors have $\|y\| < \alpha n$ and the individual in R_1 is chosen u. a. r. in line 14 ($w_2^2/2$).

Therefore,

$$\begin{aligned} p_{sel}(A(r_1, r_2)) &\geq p^2 + 2p(1 - p - p_0)\left(w_1^2 + w_1 w_2 + \frac{w_2^2}{2}\right) \\ &= p\left(p + 2(1 - p)\left(w_1^2 + w_1 w_2 + \frac{w_2^2}{2}\right)\right) \end{aligned}$$

Let $W := w_1^2 + w_1 w_2 + \frac{w_2^2}{2}$ and recall that $\Pr\left[x \in A(r_1, r_2)\right] \geq p_{sel}(A(r_1, r_2)) r_{(0)}$, therefore,

$$\Pr\left[x \in A(r_1, r_2)\right] \geq p\left(p + 2W(1-p)\right) r_{(0)}$$

Now assume that $0 < p \leq \delta'$, we will deal with $p > \delta'$ later. Thanks to the Conditions 1a, 1b or 1c, $2W - 1 > 0$. Noting that $p + 2W(1-p) = 2W - (2W-1)p$, it is clear that the minimum is attained for $p = \delta'$. Then,

$$\begin{aligned}\Pr\left[x \in A(r_1, r_2)\right] &\geq p\left(\delta' + 2W(1-\delta')\right) r_{(0)} \\ &\geq p\left(2W(1-\delta')\right) r_{(0)}\end{aligned}$$

By the assumptions on $r_{(0)}$ and $p \geq \gamma$ this is at least $(1+\delta)\gamma$.

For $p > \delta'$ we note that since $p \geq \gamma$ and $\gamma \in (0, \delta'/2]$ there is a $\delta' - \delta'/2 < \varepsilon' \leq 1 - \gamma$ such that $p = \gamma + \varepsilon'$. Then,

$$\begin{aligned}\Pr\left[x \in A(r_1, r_2)\right] \Pr\left[y \in B(s_2, s_1))\right] &\geq p r_{(0)} \\ &= \gamma\left(1 + \frac{\varepsilon'}{\gamma}\right) r_{(0)} \\ &\geq \gamma\left(1 + \frac{\delta' - \delta'/2}{\delta'/2}\right) r_{(0)} \\ &= 2\gamma r_{(0)}.\end{aligned}$$

As before, by the assumptions on $r_{(0)}$ this is at least $(1+\delta)\gamma$.

To show the statement for Condition (2) we note that since $r_2 \geq \beta n$ then the algorithm always selects a parent in $A(r_1, r_2)$ if the two individuals are sampled in R_1 and R_2 and the two competitors have $\|y\| < \alpha n$ and $\|y\| = \alpha n$. Hence,

$$\begin{aligned}p_{sel}(A(r_1, r_2)) &\geq p^2 + 2p(1 - p - p_0)\left(w_1^2 + 2w_1 w_2 + \frac{w_2^2}{2}\right) \\ &= p\left(p + 2(1-p)\left(w_1^2 + 2w_1 w_2 + \frac{w_2^2}{2}\right)\right)\end{aligned}$$

Using the same arguments as before, but using

$$r_{(0)} \geq \frac{1+\delta}{(1-\delta')2\left(w_1^2 + 2w_1 w_2 + w_2^2/2\right)},$$

we obtain the claimed results for Condition (2).

The previous sequence of levels were meant to allow Algorithm 1 build a diverse predator archive V. Once this is achieved now the same process needs to happen for the prey archive W. Since we now assume that the predator archive V is diverse, the prey population Q should move towards the maximin-optima, that is the solutions will tend towards solutions with αn 1-bits. Therefore, the levels consider the whole search space for the predator $(A(0, n))$ and for the prey the level include all solutions around the optima ($A(s_1, s_2)$ with $\alpha n - s_2 = s_1 - \alpha n$) (see Figure 5 (c)).

Note that Lemma 7 also considers levels for the predator population P that are not the whole search space. Although we do not need them in the following proofs, we leave this as this extension is not difficult to prove and may be of interest.

Lemma 7. *Let x and y be the offspring created in Lines 15 and 20 of Algorithm 1 with $k = \ell = 2$. If there are constants $\varepsilon, \delta, \delta' > 0$ such that any of the following conditions hold:*

1. *$r_1 = 0$, $r_2 \geq 0$, $s_2 \leq \alpha n$, $s_1 \geq \alpha n$ with $\alpha n - s_2 = s_1 - \alpha n$, $v_1 \geq 1 - \frac{1}{\sqrt{2}} + \varepsilon$, $v_3 \geq 1 - \frac{1}{\sqrt{2}} + \varepsilon$, $\max\left\{\frac{1+\delta}{2}, \frac{1+\delta}{1+(1-\delta')\min\left\{2\left(w_1^2+w_1w_2+\frac{w_2^2}{2}\right)-1,\left(1+\sqrt{2}\right)\varepsilon\right\}}\right\} \leq r_{(0)}^2 \leq 1$ and*
   (a) $w_1 + w_2 = 1$,
   (b) $w_1^2 > 1/2$, or
   (c) $w_1^2 \leq 1/2$ and $w_2 > \sqrt{1 - w_1^2} - w_1$
2. *$r_1 = 0$, $r_2 \geq \beta n$, $s_2 \leq \alpha n$, $s_1 \geq \alpha n$ with $\alpha n - s_2 = s_1 - \alpha n$, $v_1 \geq 1 - \frac{1}{\sqrt{2}} + \varepsilon$, $v_3 \geq 1 - \frac{1}{\sqrt{2}} + \varepsilon$, $\max\left\{\frac{1+\delta}{2}, \frac{1+\delta}{1+(1-\delta')\min\left\{2\left(w_1^2+2w_1w_2+\frac{w_2^2}{2}\right)-1,\left(1+\sqrt{2}\right)\varepsilon\right\}}\right\} \leq r_{(0)}^2 \leq 1$ and*
   (a) $w_1 + w_2 = 1$ or
   (b) $w_1^2 > 1/2$ or
   (c) $w_1^2 \leq 1/2$ and $w_2 > \sqrt{2w_1^2 + 1} - 2w_1 \geq 1 - \sqrt{2}w_1$

Then, for all $\gamma \in (0, \delta'/2]$ any population with $P \in \mathcal{X}^\lambda$ and $Q \in \mathcal{Y}^\lambda$ with $|(P \times Q) \cap (A(r_1, r_2) \times B(s_2, s_1))| \geq \gamma\lambda^2$

$$\Pr\left[x \in A(r_1, r_2)\right] \Pr\left[y \in B(s_2, s_1))\right] \geq (1 + \delta)\gamma.$$

As in Lemma 6, we note that the results of Lemma 7 can be proved for levels rotated in the search space, but we omit these here to avoid repetitive conditions.

Proof (Proof of Lemma 7). Given that the conditions in the Lemma are a combination of the conditions of Lemmas 1 and 6, this proof follows their proofs closely. We start by proving the statement for Condition (1).

Let $W := w_1^2 + w_1 w_2 + \frac{w_2^2}{2}$. From the proof of Lemma 6 the probability of creating an offspring in $A(r_1, r_2)$ is:

$$\Pr\left[x \in A(r_1, r_2)\right] \geq p\left(p + 2W(1-p)\right) r_{(0)}$$

Similarly, from the proof of Lemma 1 the probability of creating an offspring in $B(s_2, s_1)$ is:

$$\Pr\left[y \in B(s_2, s_1)\right] \geq q\left(1 + \left(1 + \sqrt{2}\right)\varepsilon(1-q)\right) r_{(0)}.$$

Then,

$$\begin{aligned}&\Pr\left[x \in A(r_1, r_2)\right]\Pr\left[y \in B(s_2, s_1))\right]\\&\geq pq\left(p + 2W(1-p)\right)\left(1 + \left(1+\sqrt{2}\right)\varepsilon(1-q)\right)r^2_{(0)}\\&= pq\left(2W - (2W-1)p)\right)\left(1 + \left(1+\sqrt{2}\right)\varepsilon - \left(1+\sqrt{2}\right)\varepsilon q)\right)r^2_{(0)}.\end{aligned}$$

Thanks to the Conditions 1a, 1b or 1c, $2W > 1$. Now assume that $0 < pq \leq \delta'$, we will deal with $pq > \delta'$ later. By Lemma 5 with $a = 2W$, $b = 1 + \left(1+\sqrt{2}\right)\varepsilon$ and $c = \delta'$ we obtain that

$$\begin{aligned}&\Pr\left[x \in A(r_1, r_2)\right]\Pr\left[y \in B(s_2, s_1))\right]\\&\geq pqr^2_{(0)}\left(1 + (1-\delta')\min\left\{2W-1, \left(1+\sqrt{2}\right)\varepsilon\right\}\right).\end{aligned}$$

By the assumptions on $r^2_{(0)}$ and $pq \geq \gamma$ this is at least $(1+\delta)\gamma$.

For $pq > \delta'$ we note that since $pq \geq \gamma$ and $\gamma \in (0, \delta'/2]$ there is an $\delta' - \delta'/2 < \varepsilon' \leq 1 - \gamma$ such that $pq = \gamma + \varepsilon'$. Then,

$$\begin{aligned}\Pr\left[x \in A(r_1, r_2)\right]\Pr\left[y \in B(s_2, s_1))\right] &\geq pqr^2_{(0)}\\&= \gamma\left(1 + \frac{\varepsilon'}{\gamma}\right)r^2_{(0)}\\&\geq \gamma\left(1 + \frac{\delta' - \delta'/2}{\delta'/2}\right)r^2_{(0)}\\&= 2\gamma r^2_{(0)}.\end{aligned}$$

As before, by the assumptions on $r^2_{(0)}$ this is at least $(1+\delta)\gamma$.

For Condition (2) we use $W := w_1^2 + 2w_1w_2 + \frac{w_2^2}{2}$ and the rest of the proof is the same as before.

Lemma 2. *Let $|V|$ and $|W|$ be the size of the archives. Then, Algorithm 3 uses at most $(|V| + \lambda)\lambda + (|W| + \lambda)\lambda$ evaluations to update the archive.*

Proof (Proof of Lemma 2). Algorithm 3 needs to check that for every solution in the current populations P there is at least one solution in V that equally ranks every solution in Q. Therefore, in the worst case the algorithm computes $g(x, y)$ for all $(x, y) \in (V \cup P) \times Q$. This amounts to $(\|V\| + \lambda)\lambda$ evaluations.

Similarly, for Q and W in the worst case it needs $(\|W\| + \lambda)\lambda$. Adding both completes the proof.

Theorem 1. *Let $\alpha, \beta \in (0, 1)$. Consider Algorithm 1 using Algorithm 3 as archive update scheme on* Bilinear$_{\alpha,\beta}$. *Define* $\mathrm{OPT} := \{(x, y) \in (\mathcal{X} \times \mathcal{Y}) \mid \|x\| = \beta n \wedge \|y\| = \alpha n\}$ *and* $T := \min\{\lambda^2 t \mid P_t \times Q_t \cap \mathrm{OPT} \neq \emptyset\}$. *Then if there are constants $\delta, \delta' > 0$ such that the probability $r_{(0)}$ of the mutation operator is at least*

$$\max\left\{\frac{8(1+\delta)}{14(1-\delta')}, \sqrt{\frac{6(1+\delta)}{6 + (1-\delta')\left(2-\sqrt{2}\right)}}\right\},$$

$r_{(1)} > 0$ *is constant and for a sufficiently large constant c,* $c \log n \leq \lambda \in \text{poly}(n)$ *then, it holds that* $\text{E}[T] = O(\lambda^4 n)$.

Proof (Proof of Theorem 1). We divide the proof in three distinct phases and we define the runtime of each phase as $T^{(i)}$ for $i \in \{1, 2, 3\}$. Phase 1 starts at the beginning of the optimisation and ends the first generation the archive V contains two solutions x_1, x_2 with $\|x_1\| < \beta n$ and $\|x_2\| > \beta n$ or the archive W contains two solutions y_1, y_2 with $\|y_1\| < \alpha n$ and $\|y_2\| > \alpha n$. Phase 2 ends when both the archive V contains two solutions x_1, x_2 with $\|x_1\| < \beta n$ and $\|x_2\| > \beta n$ and the archive W contains two solutions y_1, y_2 with $\|y_1\| < \alpha n$ and $\|y_2\| > \alpha n$. Finally Phase 3 ends when the algorithm creates a solution in OPT.

We note that if there exists a solution x with $\|x\| < \beta n$, $\|x\| = \beta n$ or $\|x\| > \beta n$ in P_t then for all $t^* \geq t$ the archive V_{t^*} contains a solution x with $\|x\| < \beta n$, $\|x\| = \beta n$ or $\|x\| > \beta n$ respectively. The same is true for the archive W_{t^*} with solutions y with $\|y\| < \alpha n$, $\|y\| = \alpha n$ or $\|y\| > \alpha n$.

We start by computing the expected runtime of Phase 1 ($T^{(1)}$). We assume that all solutions are in one quadrant of the search space, otherwise, V_0 and/or W_0 would meet the conditions to end the phase and $T^{(1)} = 1$. Due to the symmetry of the search space without loss of generality we also assume that all solutions are in the third quadrant of the search space. Then, V_0 contains exactly one solution x with $\|x\| > \beta n$ and W_0 contains a solution y with $\|y\| \leq \alpha n$. If all solutions $y \in Q_0$ have $\|y\| = \alpha n$, then in expectation we need to wait $t^* := 1 - (1 - r_{(1)})^\lambda = O(1)$ generations until a solution with $\|y\| > \alpha n$ or $\|y\| < \alpha n$ is generated by mutation and then W_{t^*} would contain a solution y with $\|y\| < \alpha n$ or $\|y\| > \alpha n$. Again by the symmetry of the search space we can assume that mutation creates y with only $\|y\| < \alpha n$.

Since, for all $0 \leq t < T^{(1)}$ W_t does not contain a solution y with $\|y\| > \alpha n$. Then $w_1 \geq 1/2$, $w_1 + w_2 = 1$, resulting in $w_1^2 + 2w_1 w_2 + w_2^2/2 \geq 7/8$ and

$$\frac{1 + \delta}{(1 - \delta')2\left(w_1^2 + 2w_1 w_2 + w_2^2/2\right)} \leq \frac{8(1 + \delta)}{14(1 - \delta')}. \tag{10}$$

We aim to use the Level-Based Theorem to show that the algorithm will find a solution in the fourth quadrant, which in turn would end Phase 1. We do not consider the event that the algorithm creates a pair of solutions $(x, y) \in P_t \times Q_t$ in the first or second quadrant. Since we want an upper bound of $\text{E}\left[T^{(1)}\right]$ and this event can only reduce $T^{(1)}$ this does not affect our computations.

Let $A^{(1)}_{i+1} := A(0, n - i)$ and $B^{(1)}_{i+1} := B(0, n)$ for $i \in [0, n(1 - \beta) + 1]$. We will use the sequence of levels $(A^{(1)}_1 \times B^{(1)}_1), \ldots, (A^{(1)}_{n(1-\beta)+2} \times B^{(1)}_{n(1-\beta)+2})$. By the assumptions on $r_{(0)}$, Equation 10 and the discussion above showing that $w_! \geq 1$ and $w_1 + w_2 = 1$ the conditions on Lemma 6 hold for all $\gamma \leq \delta'/2$ on all levels $(A^{(1)}_{i+1} \times B^{(1)}_{i+1})$ with $i \in [0, n(1 - \beta)]$. Therefore, conditions (G2a) and (G2b) of the Level-Based Theorem are met with $\gamma_0 = \delta'/2$.

A sufficient condition to create a solution in the next level $(A^{(1)}_{j+1} \times B^{(1)}_{j+1})$ is to select a solution in the current level $(A^{(1)}_j \times B^{(1)}_j)$ and flip exactly one 1-bit.

The probability of selecting a solution in the current level $(A_j^{(1)} \times B_j^{(1)})$ is at least $\gamma_0 = \delta'/2$ by definition and the probability of flipping exactly one 1-bit is at least $r_{(1)}(n-j+1)/n$. Therefore, condition (G1) is met with $z_j = \delta' r_{(1)}(n-j+1)/(2n)$. We remark that if the algorithm starts in a different quadrant this probabilities and the number of levels m might be different. To account for this let $\psi = \max\{\beta, 1-\beta, \alpha, 1-\alpha\}$. By the assumptions on λ condition (G3) of the Level-Based Theorem is met. Accounting for the initial $t^* = O(1)$ generations and $O(\lambda^2)$ evaluations that may be needed to have a solution y with $\|y\| < \alpha n$ in W_{t^*}, then for some constant $c'' > 1$

$$\begin{aligned}
\mathrm{E}\left[T^{(1)}\right] &\leq c''\lambda^2 \left((\psi n + 1)\lambda^2 + \frac{2}{\delta' r_{(1)}} \sum_{j=1}^{\psi n+1} \frac{n}{n-j+1} \right) \\
&\leq c''\lambda^2 \left(n\lambda^2 + \frac{2n}{\delta' r_{(1)}} \sum_{i=1}^{n} \frac{1}{i} \right) \\
&\leq c''\lambda^2 \left(n\lambda^2 + \frac{2n}{\delta' r_{(1)}} (1 + \ln n) \right).
\end{aligned}$$

We note that the level-based theorem in [18] assumes that each generation uses $O(\lambda)$ evaluations each generation, but this algorithm uses λ^2, this is accounted here (and in following applications of the theorem) with the λ^2 after c''.

Starting Phase 2 we assume that the archive V has at least one solution x with $x < \beta n$ and one with $x > \beta n$ and the archive W does not have a solution y with $y > \alpha n$. Hence, $v_1, v_3 \geq 1/3 \geq 1 - \frac{1}{\sqrt{2}} + \varepsilon$ for some $\varepsilon \geq \frac{1}{\sqrt{2}} - \frac{2}{3}$, $w_1 \geq 1/2$, $w_1 + w_2 = 1$ and

$$\begin{aligned}
&\frac{1+\delta}{1 + (1-\delta') \min\left\{ 2\left(w_1^2 + 2w_1w_2 + \frac{w_2^2}{2}\right) - 1, \left(1+\sqrt{2}\right)\varepsilon \right\}} \\
&\qquad \leq \frac{1+\delta}{1 + (1-\delta')\left(1+\sqrt{2}\right)\varepsilon} \\
&\qquad \leq \frac{6(1+\delta)}{6 + (1-\delta')\left(2-\sqrt{2}\right)}
\end{aligned}$$

Due to the symmetry of the search space the previous and following arguments also hold if Phase 2 starts with any other combination of archive populations.

As in Phase 1 we aim to use the Level-Based Theorem, but with the sequence of levels $(A_1^{(2)} \times B_1^{(2)}), \ldots (A_m^{(2)} \times B_m^{(2)})$ with $m = n+1$. We define $A_{n+1-i}^{(2)} := A(0, n)$ for $i \in [0, n]$, and $B_m^{(2)} := B(\alpha n{+}1, \alpha n{+}1)$ and $B_{n+1-i}^{(2)} := B(\max\{0, \alpha n - i\}, \min\{n, \alpha n + i\})$ for $i \in [1, n]$. This sequence of levels starts with the whole search space and ends with a level where $\|y\| = \alpha n + 1$.

By the assumptions on $r_{(0)}$ the conditions on Lemma 7 hold for all $\gamma \leq \delta'/2$ on all levels $(A_{n+1-i}^{(2)} \times B_{n+1-i}^{(2)})$ with $i \in [1, n]$. Therefore, conditions (G2a) and (G2b) of the Level-Based Theorem are met with $\gamma_0 = \delta'/2$.

Similar to Phase 1 we can use $z_j = \delta' r_{(1)}(n-j+1)/(2n)$ for condition (G1), and by the assumptions on λ condition (G3) of the Level-Based Theorem is met. Then for some constant $c'' > 1$

$$\mathrm{E}\left[T^{(2)}\right] \le c''\lambda^2 \left(n\lambda^2 + \frac{2n(1+\ln n)}{\delta' r_{(1)}}\right).$$

Finally for Phase 3 by definition the archive V has at least one solution x with $x < \beta n$ and one with $x > \beta n$ and the archive W has at least one solution y with $y < \alpha n$ and one with $y > \alpha n$. Hence, $v_1, v_3, w_1, w_3 \ge 1/3$, and

$$\frac{1+\delta}{1+(1-\delta')\left(1+\sqrt{2}\right)\varepsilon} \le \frac{6(1+\delta)}{6+(1-\delta')\left(2-\sqrt{2}\right)}$$

As before we will use the Level-Based Theorem with the levels

$$A^{(0)}_{2n+1-i}(0,n) \quad B^{(0)}_{2n+1-i}(\max\{0, \alpha n - i + n\}, \min\{n, \alpha n + i - n\}),$$

for $i \in [n, 2n]$ and

$$A^{(0)}_{2n+1-i}(\max\{0, \beta n - i\}, \min\{n, \beta n + i\}) \quad B^{(0)}_{2n+1-i}(\alpha n, \alpha n).$$

for $i \in [0, n-1]$.

By Lemma 1 and the assumptions on $r_{(0)}$ (G2a) and (G2b) are met. Different from previous phases the algorithm not only needs to flip the correct bit from one population, but also do not flip a bit from the other, therefore, $z_j = z_{j+n} = \delta' r_{(0)} r_{(1)}(n-j+1)/(2n)$ meets condition (G1) for all $j \in [1, m]$. By the assumptions on λ condition (G3) of the Level-Based Theorem is met. Hence, for some constant $c'' > 1$

$$\begin{aligned}
\mathrm{E}\left[T^{(3)}\right] &\le c''\lambda^2 \left((2n+1)\lambda^2 + \frac{2}{\delta' r_{(0)} r_{(1)}} \sum_{j=1}^{n+1} \frac{2n}{n-j+1}\right) \\
&\le c''\lambda^2 \left((2n+1)\lambda^2 + \frac{4n}{\delta' r_{(0)} r_{(1)}} \sum_{i=1}^{n} \frac{1}{i}\right) \\
&\le c''\lambda^2 \left((2n+1)\lambda^2 + \frac{4n(1+\log n)}{\delta' r_{(0)} r_{(1)}}\right).
\end{aligned}$$

Adding the runtime of the three phases and noting that $\lambda = \Omega(\log n)$ and $\delta, \delta', r_{(1)}, r_{(0)} > 0$ are constants we obtain

$$\mathrm{E}[T] = \mathrm{E}\left[T^{(1)}\right] + \mathrm{E}\left[T^{(2)}\right] + \mathrm{E}\left[T^{(3)}\right] = O(\lambda^4 n).$$

C.2 Lemmas and proofs of Section 6

Lemma 3. *Consider Algorithm 2 with x_1, x_2, y_1, y_2 from line 8 and x' from line 13. Let $\varepsilon \in (0, 1/2)$ be any constant. Let $A \subset \mathcal{X}$ be any subset such that*

$\forall y_1, y_2 \in \mathcal{Y}, \forall x_1 \in A, \forall x_2 \in \mathcal{X} \setminus A$, $g(x_1, y_1) \geq g(x_1, y_2)$ if and only if $g(x_2, y_1) \leq g(x_2, y_2)$, and for all $x' \in \mathcal{X} \setminus A$, $\Pr_{x \sim mut_x(x')}(x \in A) \leq \varepsilon/6$. For all $t \geq 0$, if $|P_t \cap A| \leq \frac{\lambda}{2}(1+\varepsilon)$, then $\Pr\left[|P_{t+1} \cap A| > \frac{\lambda}{2}(1+\varepsilon)\right] = e^{-\Omega(\lambda)}$.

Proof (Proof of Lemma 3). Since each new predator x in generation $t+1$ is sampled independently and identically from the same distribution, it suffices by a Chernoff bound to prove that $\Pr(x \in A) \leq (1/2)(1 + c\varepsilon)$ for some constant $c \in (0, 1)$.

First, we upper bound the probability that the selected search point x' belongs to A. The algorithm selects $x' \in A$ if both x_1 and x_2 are sampled in A, x_1 is sampled in A but x_2 and x_3 are not, or if x_2 is sampled in A but x_1 and x_3 are not. Let $\gamma := \frac{1}{\lambda}|P_t \cap A| \leq \frac{1}{2}(1+\varepsilon)$. Then

$$\begin{aligned}
\Pr[x' \in A] &= \Pr[x_1 \in A \wedge x_2 \in A] + \Pr[x_1 \in A \wedge x_2 \notin A \wedge x_3 \notin A] \\
&\quad + \Pr[x_1 \notin A \wedge x_2 \in A \wedge x_3 \notin A] \\
&= \gamma^2 + 2\gamma(1-\gamma)^2 \\
&= \gamma(2 - \gamma(3 - 2\gamma)) =: \beta(\gamma)
\end{aligned}$$

Note that since the function β is monotonically increasing in γ, and $\gamma \leq (1/2)(1+\varepsilon)$, we get

$$\begin{aligned}
\Pr[x' \in A] &\leq \frac{1}{2}(1+\varepsilon)\left(2 - \frac{1}{2}(1+\varepsilon)(3-(1+\varepsilon))\right) \\
&= \frac{1}{2}(1+\varepsilon)\left(2 - \frac{1}{2}(1+\varepsilon)(2-\varepsilon)\right) \\
&= \frac{1}{2}(1+\varepsilon)\left(1 - \frac{\varepsilon}{2} + \frac{\varepsilon^2}{2}\right) \\
&= \frac{1}{2}\left(1 + \frac{\varepsilon(1+\varepsilon^2)}{2}\right) \\
&< \frac{1}{2}\left(1 + \frac{5\varepsilon}{8}\right),
\end{aligned}$$

where the last inequality follows because $\varepsilon < 1/2$. By the law of total probability,

$$\begin{aligned}
\Pr[x \in A] &= \Pr[x' \in A]\Pr[x \in A \mid x' \in A] + \Pr[x' \notin A]\Pr[x \in A \mid x' \notin A] \\
&\leq \Pr[x' \in A] + \Pr[x \in A \mid x' \notin A] \\
&\leq \frac{1}{2}\left(1 + \frac{5\varepsilon}{8}\right) + \frac{\varepsilon}{6} \\
&= \frac{1}{2}\left(1 + \frac{23\varepsilon}{24}\right),
\end{aligned}$$

hence the statement follows by choosing the constant $c := 23/24$.

Lemma 4. *Consider Algorithm 2 with x_1, x_2, y_1, y_2 from line 8 and x' from line 13. Let $\varepsilon \in (0, 1/2)$ be any constant. Let $A \subset \mathcal{X}$ be any subset such that*

$\forall y_1, y_2 \in \mathcal{Y}, \forall x_1 \in A, \forall x_2 \in \mathcal{X} \setminus A$, $g(x_1, y_1) \geq g(x_1, y_2)$ *if and only if* $g(x_2, y_1) \leq g(x_2, y_2)$*, and for all* $x' \in A$, $\Pr_{x \sim mut_x(x')}(x \in A) \geq 1 - \frac{\varepsilon}{4}$*. For all* $t \in \mathbb{N}$*, define* $X_t := |P_t \cap A|$*. Then, for all* $t \geq 0$, $\Pr\left[X_{t+1} \geq (1 + \frac{\varepsilon}{16}) \min\{X_t, \frac{\lambda}{2}(1-\varepsilon)\} \mid X_t\right] \geq 1 - e^{-\Omega(X_t)}$.

Proof (Proof of Lemma 4). Let $\gamma := \frac{X_t}{\lambda}$, and $\gamma' := \min\{\gamma, \frac{1}{2}(1-\varepsilon)\}$. Since each new predator x in generation $t+1$ is sampled independently and identically from the same distribution, it suffices by a Chernoff bound to prove that $\Pr(x \in A) \geq \gamma'(1+\frac{\varepsilon}{8})$. More precisely, X_{t+1} is stochastically dominated by a binomially distributed random variable $X \sim \mathcal{B}(\lambda, \gamma'(1+\frac{\varepsilon}{8}))$ to which we apply Theorem 3 with parameter $\delta := \frac{\varepsilon}{16+2\varepsilon}$.

First, we lower bound the probability that the selected search point x' belongs to A, reusing the function β defined in the proof of Lemma 3. In the case $\gamma' = \gamma$, we have

$$\Pr[x' \in A] = \gamma(2 - \gamma(3 - 2\gamma))$$

noting that $\beta(\gamma)/\gamma$ is monotonically decreasing in $\gamma \in (0, 3/4)$ and $\gamma \leq \frac{1}{2}(1-\varepsilon)$

$$\begin{aligned}
&\geq \gamma\left(2 - \frac{1}{2}(1-\varepsilon)(3-(1-\varepsilon))\right) \\
&= \gamma\left(1 + \frac{\varepsilon+\varepsilon^2}{2}\right) \\
&> \gamma\left(1 + \frac{\varepsilon}{2}\right) \\
&= \gamma'\left(1 + \frac{\varepsilon}{2}\right).
\end{aligned}$$

In the case where $\gamma' < \gamma$,

$$\Pr[x' \in A] = \gamma(2 - \gamma(3 - 2\gamma))$$

noting that $\beta(\gamma)$ is monotonically increasing in γ and $\gamma > \gamma'$

$$\geq \gamma'(2 - \gamma'(3 - 2\gamma'))$$

noting that $\beta(\gamma')/\gamma'$ is monotonically decreasing in γ' gives as above

$$\begin{aligned}
&\geq \gamma'\left(2 - \frac{1}{2}(1-\varepsilon)(3-(1-\varepsilon)\right) \\
&= \gamma'\left(1 + \frac{\varepsilon+\varepsilon^2}{2}\right) \\
&> \gamma'\left(1 + \frac{\varepsilon}{2}\right)
\end{aligned}$$

By the law of total probability,

$$\begin{aligned}\Pr[x \in A] &= \Pr[x' \in A]\Pr[x \in A \mid x' \in A] + \Pr[x' \notin A]\Pr[x \in A \mid x' \notin A] \\ &\geq \Pr[x' \in A]\Pr[x \in A \mid x' \in A] \\ &\geq \gamma'\left(1+\frac{\varepsilon}{2}\right)\left(1-\frac{\varepsilon}{4}\right) \\ &= \gamma'\left(1+\frac{\varepsilon}{4}\left(1-\frac{\varepsilon}{2}\right)\right) \\ &\geq \gamma'\left(1+\frac{\varepsilon}{8}\right).\end{aligned}$$

For ease of notation, we will introduce the following random variables $X_t := |P_t \cap R_0|$ and $Y_t := |Q_t \cap S_0|$.

Phase 1 Phase 1 starts in generation t_0 with the assumptions in Eq. (2). (Recall that a symmetry argument can be applied if this assumption does not hold). In a successful Phase 1, the populations satisfy predicate $\mathcal{E}_1$ from generation t_0 until at least generation t_7. Informally, this means that not much more than a quarter of the predator-prey pairs belong to the first quadrant $R_0 \times S_0$.

Lemma 8. *Under Assumption 2, Phase 1 with* $\tau_1 = 0$ *and predicate* $\mathcal{E}_1$ *is successful with probability at least* $1 - 2\tau e^{-\Omega(\lambda)}$.

Proof (Proof of Lemma 8). If $X_t \leq \frac{\lambda}{2}(1+\varepsilon)$, then by Lemma 3 with parameter $A = R_0$, the probability that for all $t \in [t_2, t_7]$, it holds $X_{t+1} \leq \frac{\lambda}{2}(1+\varepsilon)$ is at least $1-e^{-\Omega(\lambda)}$. By symmetry of the problem and the algorithm, if $Y_t \leq \frac{\lambda}{2}(1+\varepsilon)$, then the probability that $Y_{t+1} \leq \frac{\lambda}{2}(1+\varepsilon)$ is at least $1-e^{-\Omega(\lambda)}$. Furthermore, for any $x' \in \mathcal{X} \setminus A$,

$$\begin{aligned}\Pr_{x\sim \mathrm{mut}_x(x')}(x \in A) &\leq 1 - \Pr_{x\sim \mathrm{mut}_x(x')}(x = x') \\ &= 1 - r_{(0)} \\ &\leq \varepsilon/6,\end{aligned}$$

where the last inequality follows from Assumption 2. The lemma now follows by a union bound over $t_7 - t_2 \leq \tau$ generations.

Phase 2 Assuming no failure in Phase 1, Phase 2 starts at generation t_2 with at most $\frac{\lambda}{2}(1+\varepsilon)$ predators in R_0, and at most $\frac{\lambda}{2}(1+\varepsilon)$ prey in S_0 (predicate $\mathcal{E}_1$). After generation t_2 of a successful Phase 2, at least $\gamma_0\lambda$ predators belong to region R_0 (predicate $\mathcal{E}_2$).

To analyse the success probability of Phase 2, we apply Theorem 4 with the $m = (1-\beta)n + 1$ levels defined $\forall j \in [m]$ by

$$A_j := \begin{cases} \mathcal{X} & \text{if } j = 0, \text{ and} \\ R_0 \cup R_1(j) & \text{otherwise,} \end{cases} \qquad B_j := \mathcal{Y} \tag{11}$$

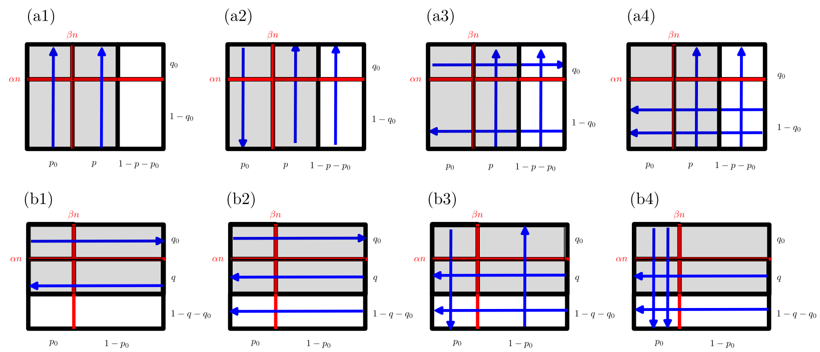

Fig. 6: (a1)–(a4): Four ways of selecting a predator in $R_0 \cup R_1$ during Phase 2. (b1)–(b4): Four ways of selecting a prey in $S_0 \cup S_1$ during Phase 4.

These levels are illustrated in Figure 6, (a1)–(a4), and satisfy the following inclusions $R_0 = A_m \subset A_{m-1} \subset \cdots \subset A_{j+1} \subset A_j \subset \cdots \subset A_0 = \mathcal{X}$.

Lemma 9. *Let $\varepsilon, \gamma_0, \delta \in (0,1)$ be as in Definition 3. If $q_0 \leq (1/2)(1+\varepsilon)$ and $p_0 + p = \gamma \leq \gamma_0$ and $\gamma_0 + \varepsilon < \frac{1-2\delta}{3}$, then $p_{\mathrm{sel}}(R_0 \cup R_1) \geq \gamma(1+\delta)$.*

Proof (Proof of Lemma 9). Following cases (a1)–(a4) in Figure 6, the algorithm selects a predator in region $R_0 \cup R_1$ in the if: $x_1, x_2 \in R_0 \cup R_1$ (a1), $x_1 \in R_0$, $x_2 \in R_2$ and $x_3 \in R_1$ (a2), $x_1 \in R_1$, $x_2 \in R_2$, $y_1 \in S_0$ and $y_2 \notin S_0$ (a3), and $x_1 \in R_1$, $x_2 \in R_2$, $y_1, y_2 \notin S_0$ (a4). Additionally, the cases (a2)–(a4) have alternatives where x_2 is exchanged for x_1 or x_3 and y_1 is exchanged for y_2. Adding all these cases we obtain,

$$\begin{aligned}
p_{\mathrm{sel}}(R_0 \cup R_1) &= (p_0+p)^2 + 2p_0(1-p-p_0)(1-p_0) + 2p(1-p-p_0)2q_0(1-q_0) + 2p(1-p-p_0)(1-q_0)^2 \\
&\geq \gamma^2 + 2p_0(1-\gamma)^2 + 2p(1-\gamma)(1-q_0)(2q_0+1-q_0) \\
&= \gamma^2 + 2p_0(1-\gamma)^2 + 2p(1-\gamma)(1-q_0^2) \\
&\geq \gamma^2 + 2p_0(1-\gamma)^2 + 2p(1-\gamma)\left(1 - \frac{1}{4}(1+\varepsilon)^2\right) \\
&= \gamma^2 + 2p_0(1-\gamma)^2 + 2p(1-\gamma)\left(\frac{3}{4} - \frac{\varepsilon}{2} - \frac{\varepsilon^2}{4}\right) \\
&> \gamma^2 + 2(1-\gamma)\left(p_0(1-\gamma) + p\frac{3}{4}(1-\varepsilon)\right) \\
&> \gamma^2 + 2(1-\gamma_0)\frac{3}{4}(1-\varepsilon)(p_0+p) \\
&= \gamma\left(\gamma + \frac{3}{2}(1-\gamma_0)(1-\varepsilon)\right) \\
&> \gamma\left(1 + \frac{1}{2} - \frac{3}{2}(\gamma_0+\varepsilon)\right) \\
&> \gamma(1+\delta).
\end{aligned}$$

The next lemma provides an upper bound on the the expected time to satisfy predicate $\mathcal{E}_2$.

Lemma 10. *Under Assumptions 1 and 2, if Phase 1 is successful, then Phase 2 with* $\tau_2 = O\left(n\lambda^2(1-\beta) + n\ln(1/\beta)\right)$ *is successful with probability at least* $9/10 - \tau e^{-\Omega(\lambda)}$.

Proof (Proof of Lemma 10). In the following, we let T denote the runtime until condition $\mathcal{E}_2$ is satisfied assuming that Phase 1 was successful. We apply Theorem 4 with the $m = n(1-\beta)+1$ levels defined in (11).

We first verify condition (G2a). Since $B_{j+1} = \mathcal{Y}$, $\Pr[y \in \mathcal{Y}] = 1$. Now, for any $j \in [0..m-1]$, suppose that $\|(P_t \times Q_t) \cap (A_{j+1} \times B_{j+1})\| \geq \gamma\lambda^2$. Then, $p_0 + p = |P_t \cap A_{j+1}|/\lambda^2 = \gamma$, and by Lemma 9, we have $\Pr[x' \in A_{j+1}] \geq \gamma(1+\delta)$. Additionally by by Assumption 2 $\Pr[x \in A_{j+1} \mid x' \in A_{j+1}] \geq r_{(0)} \geq (1-\delta/2)$. Then,

$$\begin{aligned}
&\Pr[x \in A_{j+1}] \Pr[y \in B_{j+1}] \\
&\geq \Pr[x' \in A_{j+1}] \Pr[x \in A_{j+1} \mid x' \in A_{j+1}] \Pr[y \in \mathcal{Y}] \\
&\geq \gamma(1+\delta) r_{(0)} \\
&\geq \gamma(1+\delta)(1-\delta/2) \\
&\geq \gamma\left(1 + \frac{\delta}{2}(1-\delta)\right).
\end{aligned}$$

Hence, condition (G2a) is satisfied. Condition (G2b) can be shown following the same steps but using A_j, B_j and γ_0.

We now prove condition (G1). Since $B_{j+1} = \mathcal{Y}$ for all j, we have

$$\Pr[x \in A_{j+1}] \Pr[y \in B_{j+1}] = \Pr[x \in A_{j+1}].$$

It therefore remains to compute the probability that an offspring predator x belongs to level A_{j+1} assuming we already have at least $\gamma_0\lambda$ predators in A_j. If the selected predator x' is in $A_j \setminus A_{j+1}$, then x has exactly j 0-bits, and it suffices to flip one 1-bit and no other bits to produce an offspring in A_{j+1}. If x' already belongs to A_{j+1}, then it suffices to flip no bits, we pessimistically assume this does not happen. Hence, in overall, we have

$$\begin{aligned}
\Pr[x \in A_{j+1}] &\geq \Pr[x' \in A_j] \Pr[x \in A_{j+1} \mid x' \in A_j] \\
&\geq \gamma_0(1+\delta)\frac{(n-j)r_{(1)}}{n} =: z_j.
\end{aligned}$$

Condition (G1) is therefore satisfied for the parameters $z_j = \Theta(1 - j/n)$.

Finally, condition (G3) is satisfied by choosing $\lambda \geq c\log(n)$ for a sufficiently large constant c.

By Theorem 4, starting from any initial configuration, for some sufficiently large constant c'' and $r = 1/10$, the time (in terms of function evaluations) until

$|P_t \cap R_0| \geq \gamma_0\lambda$ satisfies $\Pr[T \geq \lambda\tau_2] < 1/10$ with

$$\begin{aligned}
\tau_2 &= O(\lambda^2 m + \sum_{j=0}^{m-1} 1/z_j) \\
&= O\left(\lambda^2 n(1-\beta) + \sum_{j=0}^{n(1-\beta)-1} \frac{n}{n-j}\right) \\
&= O\left(\lambda^2 n(1-\beta) + n\left(\sum_{j=1}^{n} \frac{1}{j} - \sum_{j=1}^{\beta n} \frac{1}{j}\right)\right) \\
&= O\left(\lambda^2 n(1-\beta) + n \ln\left(\frac{n}{\beta n}\right)\right) \\
&= O\left(\lambda^2 n(1-\beta) + n\ln(1/\beta)\right).
\end{aligned}$$

Note that for all $x' \in A$, by Assumption 2,

$$\begin{aligned}
\Pr_{x\sim \mathrm{mut}_x(x')}(x \in A) &\geq \Pr_{x\sim \mathrm{mut}_x(x')}(x = x') \\
&= r_{(0)} \\
&\geq 1 - \varepsilon/6 \\
&> 1 - \varepsilon/4,
\end{aligned}$$

hence the assumption of Lemma 4 is satisfied.

Therefore, if $X_{t_2} \geq \gamma_0\lambda$, then by Lemma 4 and a union bound, the probability that for all $t \in [t_2, t_7]$, it holds $X_t \geq \gamma_0\lambda$ is at least $1 - \tau e^{-\Omega(\lambda)}$. The statement now follows by a union bound.

Phase 3 After the next phase, we need to assure that at least $\frac{\lambda}{2}(1-\varepsilon)$ belong to region R_0 (predicate $\mathcal{E}_3$). Since predators in R_0 rank prey differently than predators in $R_1 \cup R_2$, the diversity mechanism in the algorithm ensures that R_0-predators quickly expand once they are discovered, as shown in Lemma 4.

Lemma 11. *If Phase 2 is successful, then Phase 3 with* $\tau_3 = O(1)$ *is successful with probability* $1 - \tau e^{-\Omega(\lambda)}$.

Proof (Proof of Lemma 11). Note first that since γ_0 and ε are constants, there exists a $\tau_3 = O(1)$ such that $X_{t_2}(1 + \frac{\varepsilon}{16})^{\tau_3} \geq \frac{\lambda}{2}(1-\varepsilon)$. Hence, by Lemma 4 and a union bound, for all $t \in [t_3, t_7]$, $X_t \geq \frac{\lambda}{2}(1-\varepsilon)$ with probability at least $1 - \tau e^{-\Omega(\lambda)}$.

Phase 4 If Phase 3 is successful, then Phase 4 starts with $X_{t_4} \geq \frac{\lambda}{2}(1-\varepsilon)$. After generation t_4 of a successful Phase 4, at least $\gamma_0\lambda$ prey belong to region S_0.

Note that the analysis of Phase 4 is analogous to the analysis of Phase 2. In Phase 4, we will consider the following levels.

$$A_j := \mathcal{X} \qquad B_j := \begin{cases} \mathcal{Y} & \text{if } j = 0, \text{ and} \\ S_0 \cup S_1(j) & \text{otherwise,} \end{cases} \tag{12}$$

The levels are illustrated in Figure 6 (b1)–(b4). Note the following inclusion in the levels $S_0 = B_m \subset B_{m-1} \subset \cdots \subset B_{j+1} \subset B_j \subset \cdots \subset B_0 = \mathcal{Y}$.

Lemma 12. *Let $\varepsilon, \gamma_0, \delta \in (0,1)$ be as in Definition 3. If $(1/2)(1-\varepsilon) \leq p_0 \leq (1/2)(1+\varepsilon)$ and $q_0 + q = \gamma$, then for sufficiently small γ_0, there exists a $\delta > 0$ such that $p_{\mathrm{sel}}(S_0 \cup S_1) \geq \gamma(1+\delta)$.*

Proof (Proof of Lemma 12). Following cases (b1)–(b4) in Figure 6, the algorithm selects a prey in region $S_0 \cup S_1$ if: $y_1, y_2 \in S_0 \cup S_1$ (b1), $y_1 \in S_0$, $y_2 \in S_2$ and $y_3 \in S_3$ (b2), $y_1 \in S_1$, $y_2 \in S_2$, $x_1 \in R_0$ and $x_2 \notin R_0$ (b3), and $y_1 \in S_1$, $Y_2 \in S_2$, $x_1, x_2 \notin R_0$ (b4). Additionally, the cases (b2)–(b4) have alternatives where y_2 is exchanged for y_1 or y_3 and x_1 is exchanged for x_2. Adding all these cases we obtain

$$\begin{aligned} p_{\mathrm{sel}}(S_0 \cup S_1) &\geq (q_0+q)^2 + 2q_0(1-q-q_0)(1-q_0) + \\ &\qquad 2(1-q-q_0)qp_0(2-2p_0+p_0) \\ &\geq \gamma^2 + 2q_0(1-\gamma)^2 + 2q(1-\gamma)\frac{1}{2}(1-\varepsilon)\left(2 - \frac{1}{2}(1+\varepsilon)\right) \end{aligned}$$

for sufficiently small ε

$$\begin{aligned} &> \gamma^2 + 2q_0(1-\gamma)^2 + \frac{5}{4}q(1-\gamma) \\ &> \gamma^2 + \frac{5}{4}(1-\gamma)^2(q_0+q)) \\ &= \gamma\left(\frac{5}{4} - \frac{3}{2}\gamma + \gamma^2\right) \\ &> \gamma\left(\frac{5}{4} - \frac{3}{2}\gamma_0\right) \\ &> \gamma(1+\delta) \end{aligned}$$

where the last inequality holds for sufficiently small γ_0.

Lemma 13. *Under Assumptions 1 and 2, if Phase 3 is successful, then Phase 4 with $\tau_4 = O\left(n(\lambda^2(1-\alpha) + \ln(1/\alpha)\right)$ is successful with probability $9/10 - \tau e^{-\Omega(\lambda)}$.*

Proof (Proof of Lemma 13). By using the symmetry between R_0 and S_0, the proof is analogous to the proof of Lemma 10, but uses Lemma 12 instead of Lemma 9.

Phase 5 Phase 5 is analogous to Phase 3. If Phase 4 is successful, the algorithm has obtained at least $\gamma_0\lambda$ prey in S_0, and it is straightforward to prove that the algorithm quickly acquires atlemmaallend least $\frac{\lambda}{2}(1-\varepsilon)$ prey in S_0 (predicate $\mathcal{E}_5$).

Lemma 14. *If Phase 4 is successful, then Phase 5 with $\tau_3 = O(1)$ is successful with probability $1 - \tau e^{-\Omega(\lambda)}$.*

Proof (Proof of Lemma 14). The proof is analogous to the proof of Lemma 11.

Phase 6 If Phase 5 is successful, then Phase 6 starts with $X_t \geq \frac{\lambda}{2}(1-\varepsilon)$ and $Y_t \geq \frac{\lambda}{2}(1-\varepsilon)$. Informally, after generation t_6, $\gamma_0\lambda$ predators correspond to the optimum for the predators.

To analyse the success probability of Phase 6, we apply Lemma 4 with the levels $A_j := R_1(j)$, $B_j := S_1(0)$.

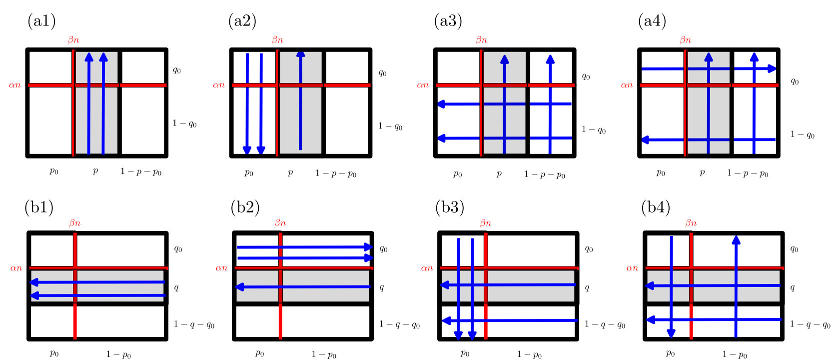

Fig. 7: (a1)–(a4): Four ways of selecting a predator in $R_0 \cup R_1$ during Phases 6 and 7. (b1)–(b4): Four ways of selecting a prey in $S_0 \cup S_1$ during Phase 6 and 7.

Lemma 15. *Let $\varepsilon, \gamma_0, \delta \in (0,1)$ be as in Definition 3. If $(1/2)(1-\varepsilon) \leq p_0 \leq (1/2)(1+\varepsilon)$ and $(1/2)(1-\varepsilon) \leq q_0 \leq (1/2)(1+\varepsilon)$, and $pq \leq \gamma_0$, then*

$$\frac{p_{\text{sel}}(R_1)}{p} \geq \frac{1}{4}(5 - 2p - 9\varepsilon), \text{ and}$$
$$\frac{p_{\text{sel}}(S_1)}{q} \geq \frac{5}{4} - \frac{q}{2} - \frac{11\varepsilon}{4}.$$

Furthermore, if ε and γ_0 are sufficiently small constants, there exists a constant $\delta > 0$ such that

$$\frac{p_{\text{sel}}(R_1)}{p}\frac{p_{\text{sel}}(S_1)}{q} \geq 1 + \delta.$$

Proof (Proof of Lemma 15). As illustrated in cases (a1)–(a4) in Figure 7, the algorithm selects a search point $x' \in R_1$ in four cases (and their permutations with respect to x_1, x_2, x_3, y_1, y_2): $x_1, x_2 \in R_1$ (a1), $x_1 \in R_1$, $x_2 \in R_0$ and $x_3 \in R_0$ (a2), $x_1 \in R_1$, $x_2 \in R_2$, $y_1, y_2 \notin S_0$ (a3), and $x_1 \in R_1$, $x_2 \in R_2$, $y_1 \in S_0$ and $y_2 \notin S_0$ (a4) having the following probabilities,

$$\begin{aligned} p_{\text{sel}}(R_1) &= p^2 + 2pp_0^2 + 2p(1-p-p_0)(1-q_0)^2 + 2p(1-p-p_0)2q_0(1-q_0) \\ &= p(p + 2p_0^2 + 2(1-p-p_0)(1-q_0)(1-q_0+2q_0)) \\ &= p(p + 2p_0^2 + 2(1-p-p_0)(1-q_0^2)) \end{aligned}$$

Using the assumptions on p_0 and q_0, we obtain

$$\begin{aligned} \frac{p_{\text{sel}}(R_1)}{p} &\geq p + \frac{2}{4}(1-\varepsilon)^2 + 2\left(1-p-\frac{1}{2}(1+\varepsilon)\right)\left(1-\frac{1}{4}(1+\varepsilon)^2\right) \\ &> p + \frac{1}{2}(1-2\varepsilon) + 2\left(\frac{1}{2}-p-\frac{\varepsilon}{2}\right)\left(\frac{3}{4}-\frac{\varepsilon}{2}-\frac{\varepsilon^2}{4}\right) \\ &= \frac{5}{4} - \frac{p}{2} + \frac{p\varepsilon^2}{2} + p\varepsilon + \frac{\varepsilon^3}{4} + \frac{\varepsilon^2}{4} - \frac{9\varepsilon}{4} \\ &> \frac{5}{4} - \frac{p}{2} - \frac{9\varepsilon}{4} \\ &> \frac{5}{4} - \frac{p}{2} - 3\varepsilon. \end{aligned}$$

Analogously to $p_{\text{sel}}(R_1)$, we have

$$\begin{aligned} p_{\text{sel}}(S_1) &= q^2 + 2qq_0^2 + 2q(1-q-q_0)p_0^2 + 2q(1-q-q_0)2p_0(1-p_0) \\ &= q(q + 2q_0^2 + 2(1-q-q_0)p_0(p_0 + 2(1-p_0)) \\ &= q(q + 2q_0^2 + 2(1-q-q_0)p_0(2-p_0)) \end{aligned}$$

Applying the assumptions on p_0 and q_0, we obtain

$$\begin{aligned} \frac{p_{\text{sel}}(S_1)}{q} &\geq q + \frac{2}{4}(1-\varepsilon)^2 + 2\left(1-q-\frac{1}{2}(1+\varepsilon)\right)\frac{1}{2}(1-\varepsilon)\left(2-\frac{1}{2}(1+\varepsilon)\right) \\ &\geq q + \frac{1}{2}(1-2\varepsilon) + \left(\frac{1}{2}-q-\frac{\varepsilon}{2}\right)(1-\varepsilon)\left(\frac{3}{2}-\frac{\varepsilon}{2}\right) \\ &= \frac{5}{4} - \frac{q}{2} - \frac{q\varepsilon^2}{2} + 2q\varepsilon - \frac{\varepsilon^3}{4} + \frac{5\varepsilon^2}{4} - \frac{11\varepsilon}{4} \\ &> \frac{5}{4} - \frac{q}{2} - \frac{11\varepsilon}{4} \\ &> \frac{5}{4} - \frac{q}{2} - 3\varepsilon. \end{aligned}$$

For the rest of the proof, we can assume without loss of generality that $q \leq p$, since they contribute equally to $\frac{p_{\text{sel}}(R_1)}{p}$ and $\frac{p_{\text{sel}}(S_1)}{q}$ respectively. We therefore have $q^2 \leq pq \leq \gamma_0$ so $q \leq \sqrt{\gamma_0}$. Furthermore, since $p_0 \geq \frac{1}{2}(1-\varepsilon)$, we must also have $p \leq \frac{1}{2}(1+\varepsilon)$. This now gives

$$\begin{aligned}
\frac{p_{\text{sel}}(R_1)}{p}\frac{p_{\text{sel}}(S_1)}{q} &\geq \left(\frac{5}{4} - \frac{p}{2} - 3\varepsilon\right)\left(\frac{5}{4} - \frac{q}{2} - 3\varepsilon\right) \\
&\geq \left(\frac{5}{4} - \frac{1}{4}(1+\varepsilon) - 3\varepsilon\right)\left(\frac{5}{4} - \frac{\sqrt{\gamma_0}}{2} - 3\varepsilon\right) \\
&= \left(1 - \frac{13\varepsilon}{4}\right)\left(\frac{5}{4} - \frac{\sqrt{\gamma_0}}{2} - 3\varepsilon\right) \\
&\geq 1 + \delta.
\end{aligned}$$

The last inequality holds for a suitable choice of ε and γ_0

Lemma 16. *Under Assumptions 1 and 2, if Phase 5 is successful, then Phase 6 with* $\tau_6 = O(n\lambda^2(1-\beta) + n\ln(1/\beta))$ *is successful with probability* 9/10.

Proof (Proof of Lemma 16). The proof uses Theorem 4, analogously to Lemma 10. In particular, Conditions (G2a) and (G2b) are satisfied due to Lemma 15. Given that the algorithm is at the current level j Condition (G1) can be satisfied for the parameters

$$z_j \geq \gamma_0(1+\delta)\left(\frac{n-j}{n}\right) r_{(0)} r_{(1)},$$

corresponding to the probability of selecting a predator-prey pair (x', y') in $R_1(j) \times S_1(0)$ (which probability can be bounded using Lemma 15), not flipping any bit in the prey, and flipping at most one 1-bit in the predator.

Phase 7 If Phase 6 is successful, then the final Phase 7 begins with $|P_t \cap R_1(n(1-\beta)-1)| \geq \gamma_0\lambda$. In a successful Phase 7, the optimum is found in generation t_7.

To obtain a lower bound on the success probability in Phase 7, we apply Theorem 4 with the levels

$$A_j := R_1(n(1-\beta)-1) \qquad B_j := S_1(j). \tag{13}$$

Lemma 17. *Under Assumptions 1 and 2, if Phase 6 is successful, then Phase 7 with* $\tau_7 = O\left((1-\alpha)n\lambda^2 + n\ln(1/\alpha)\right)$ *is successful with probability* 9/10.

Proof (Proof of Lemma 17). We apply Theorem 4 with the $m = n(1-\alpha)+1$ levels defined in (13).

We first verify condition (G2a). For any $j \in [0..m-1]$, suppose that $(P_t \times Q_t) \cap (A_{j+1} \times B_{j+1}) \geq \gamma\lambda^2$. Then, $pq = \gamma$, and by Lemma 15, we have

$$\begin{aligned}
&\Pr[x \in A_{\geq j+1}] \Pr[y \in B_{\geq j+1}] \\
&\geq \Pr[x' \in A_{\geq j+1}] \Pr[x \in A_{\geq j+1} \mid x' \in A_{\geq j+1}] \\
&\quad \cdot \Pr[y' \in B_{\geq j+1}] \Pr[y \in B_{\geq j+1} \mid y' \in B_{\geq j+1}] \\
&\geq \gamma(1+\delta) r_{(0)}^2 \\
&\geq \gamma(1+\delta)(1-\delta/2) \\
&= \gamma\left(1 + \frac{\delta}{2}(1-\delta)\right).
\end{aligned}$$

Hence, condition (G2a) is satisfied. Condition (G2b) can be shown similarly.

We now prove condition (G1). Assume that there exist at least $\gamma_0\lambda$ pairs in $A_j \times B_j$ as assumed by Theorem 4. We need to lower bound the probabilities of producing a predator in A_{j+1} and a prey in B_{j+1}. Since $A_{j+1} = R_1(n(1-\beta)-1)$ for all j, it suffices for the first probability to estimate the probability of selecting a predator in A_j and not flipping any bits. The probability of this event is at least

$$\begin{aligned}
\Pr[x \in A_{j+1}] &\geq \Pr\left[x' \in A_j^{(1)}\right] \Pr\left[x \in A_{j+1}^{(1)} \mid x' \in A_j^{(1)}\right] \\
&= p_{\text{sel}}(R_1) r_{(0)}.
\end{aligned}$$

We now compute the probability that an offspring prey y belongs to level B_{j+1}. If the selected prey y' is in $B_j \setminus B_{j+1}$, then y has exactly $n-j$ 0-bits, and it suffices to flip one 0-bit and no other bits to produce an offspring in B_{j+1}. If y' already belongs to B_{j+1}, then it suffices to flip no bits. Hence, in overall, we have

$$\begin{aligned}
\Pr[y \in B_{j+1}] &\geq p_{\text{sel}}(S_1) \Pr[x \in B_{j+1} \mid x' \in B_j] \\
&\geq p_{\text{sel}}(S_1)(n-j) r_{(1)}.
\end{aligned}$$

In overall, we have

$$\Pr[x \in A_{j+1}] \Pr[y \in B_{j+1}] \geq p_{\text{sel}}(R_1) p_{\text{sel}}(S_1)(n-j) r_{(0)} r_{(1)}$$

by Lemma 15

$$\geq (1+\delta)\gamma_0 \left(\frac{n-j}{n}\right) r_{(0)} r_{(1)} := z_j.$$

Condition (G1) is therefore satisfied for the parameters $z_j = \Theta(1 - j/n)$.

Finally, condition (G3) is satisfied by choosing $\lambda \geq c\log(n)$ for a sufficiently large constant c.

By Theorem 4 with $r = 10$ and a sufficiently large constant c'', the time T (in function evaluations) until $P_t \times Q_t \cap (S_1(n(1-\beta)) \times R_1(\alpha n)) \neq \emptyset$ satisfies

$\Pr[T \geq \lambda\tau_7] < 1/10$ where

$$\begin{aligned}\tau_7 &= O\left(\lambda^2 m + \sum_{j=0}^{m-1} 1/z_j\right) \\ &= O\left(\lambda^2 m + \sum_{j=0}^{n(1-\alpha)} \frac{n}{n-j}\right) \\ &= O\left((1-\alpha)n\lambda^2 + n\ln(1/\alpha)\right)\end{aligned}$$

Theorem 2. *For all $\alpha n, \beta n \notin \mathbb{Z}$, Algorithm 2 with parameter satisfying Assumption 2 has expected runtime $O(n\lambda^3 - n\lambda\ln(\alpha\beta(1-\beta)(1-\alpha)))$ on $\text{BILINEAR}_{\alpha,\beta}$.*

Proof (Proof of Theorem 2). Under Assumption 1, the statement follows from Lemmas 8, 10, 11, 13, 14, 16, and 17. In particular, each era lasts $\tau = \sum_{i=1}^{7} \tau_i = O((2-\alpha-\beta)n\lambda^2 + n\ln(1/(\alpha\beta)))$ generations and is successful with probability $\Omega(1)$. Hence, a successful era occurs in expectation after $O(1)$ eras, or equivalently, after $O(\lambda\tau)$ evaluations of the payoff function.

We now consider the case that Assumption 1 does not hold. In the case of the "rotated" assumption $|P_{t_0} \cap R_0| > \frac{\lambda}{2}(1+\varepsilon)$ and $|Q_{t_0} \cap S_0| > \frac{\lambda}{2}(1+\varepsilon)$, it suffices to consider the predicates

$$\begin{aligned}\mathcal{E}_2(P,Q) &:= \gamma_0\lambda \leq |P \cap (R_1 \cup R_2)| \\ \mathcal{E}_3(P,Q) &:= \frac{\lambda}{2}(1-\varepsilon) \leq |P \cap (R_1 \cup R_2)| \\ \mathcal{E}_4(P,Q) &:= \gamma_0\lambda \leq |P \cap (S_1 \cup S_2)| \\ \mathcal{E}_5(P,Q) &:= \frac{\lambda}{2}(1-\varepsilon) \leq |P \cap (S_1 \cup S_2)| \\ \mathcal{E}_6(P,Q) &:= \frac{\lambda}{2}(1-\varepsilon) \leq |Q \cap S_1(\alpha n - 1)| \\ \mathcal{E}_7(P,Q) &:= \exists x \in P, \exists y \in Q, \|x\| = \beta n \wedge \|y\| = \alpha n\end{aligned}$$

An equivalent analysis shows that the era in this case lasts $\tau = O((\beta + 1 - \alpha)n\lambda^2 + n\ln(1/(\alpha(1-\beta)))$ generations and is successful with probability $\Omega(1)$. The overall runtime now follows by considering all rotations of Assumption 1.