Analysis of local-search heuristics for Mastermind with n colors

Here we prove tight lower bounds and upper bounds for RLS, and almost tight lower bounds and upper bounds for $(1+1)$-EA.

Notation.

- $\log$ denotes the logarithm in base e.
- $[n] \overset{\text{def}}{=} \{1, \ldots, n\}$
- $H_n \overset{\text{def}}{=} \sum_{i=1}^{n} \frac{1}{i}$

Definition (MM_n)**.** As all our algorithms will be unbiased, we consider the objective function $f : [n]^n \to \mathbf{R}_+$ defined by:

$$f(x) \overset{\text{def}}{=} \sum_{i=1}^{n} \mathbf{1}_{x_i=0}$$

Remark. For RLS, we initialize $x \in_R [n]^n$ and at each step we choose an index $i \in_R [n]$ and a shift $s \in_R [n-1]$.

Theorem.

$$\mathbf{E}[\mathcal{T}(RLS, MM_n)] \underset{n\to\infty}{\sim} n^2 \log n$$

Proof. Let $x^0, \ldots x^t$ be the successive queries of RLS. Let $T_i \overset{\text{def}}{=} \min\{t \mid f(x^t) \geq i\}$. Note that for all $i \geq i_0$, the variable $T_{i+1} - T_i$ conditioned on $f(x^0) = i_0$ follows a geometric distribution with parameter $\frac{n-i}{n}\frac{1}{n-1}$. So:

$$\mathbf{E}[T_{i+1} - T_i \mid f(x^0) = i_0] = \frac{n(n-1)}{n-i}$$

And it follows that:

$$\mathbf{E}[T_n \mid f(x^0) = i_0] = n(n-1)\sum_{i=i_0}^{n-1} \frac{1}{n-i} = n(n-1)H_{n-i_0}$$

If $c_n \overset{\text{def}}{=} \sqrt{n \log n}$, by additive Chernoff bounds:

$$\mathbf{P}[f(x^0) > 1 + c_n] \leq \exp(-2\log n) = \frac{1}{n^2}$$

So:

$$\mathbf{E}[T_n] = n(n-1) \sum_{i_0 \leq 1+c_n} H_{n-i_0} \mathbf{P}[f(x^0) = i_0] + o(1)$$

Finally:

$$\left(1 - \frac{1}{n^2}\right) \log(n-1-c_n) \leq \sum_{i_0 \leq 1+c_n} H_{n-i_0} \mathbf{P}[f(x^0) = i_0] \leq 1 + \log n$$

Hence $\mathbf{E}[T_n] \underset{n\to\infty}{\sim} n^2 \log n$. □

Remark. For $(1+1)$-EA, we consider $p = \frac{1}{n}$.

Theorem.

$$\mathbf{E}[\mathcal{T}((1+1)\text{-}EA, MM_n)] \leq en^2(\log n + 1)$$

Proof. We apply the fitness local method. Let x be the current state of the variable and y the transformed one.

$$\mathbf{P}[f(y) > i \mid f(x) = i] \geq (n-i) \left(1 - \frac{1}{n}\right)^{n-1} \frac{1}{n^2} \geq \frac{n-i}{en^2}$$

So:

$$\mathbf{E}[\mathcal{T}((1+1)\text{-EA}, \text{MM}_n)] \leq en^2 H_n \leq en^2(\log n + 1)$$

□

Theorem.

$$\mathbf{E}[\mathcal{T}((1+1)\text{-}EA, MM_n)] \geq n^2 \log n + o(n^2 \log n)$$

Proof. Let $X_{j,t}$ be the indicator of the event "the j-th position is incorrect in x^0 and zero has never been drawn out for this position in the first t iterations". Let:

$$t_n \overset{\text{def}}{=} \left(1 - \frac{\log \log n}{\log n}\right) (n^2 - 1) \log n$$

Then at this time:

$$\mathbf{P}[X_{j,t_n} = 1] = \left(1 - \frac{1}{n}\right) \left(1 - \frac{1}{n^2}\right)^{\left(1 - \frac{\log \log n}{\log n}\right)(n^2-1) \log n} \geq \frac{\log n}{2n}$$

We deduce:

$$\mathbf{E}\left[\sum_{j=1}^{n} X_{j,t_n}\right] \geq \frac{\log n}{2}$$

And by multiplicative Chernoff bounds:

$$\mathbf{P}\left[\sum_{j=1}^{n} X_{j,t_n} \leq \frac{\log n}{4}\right] \leq \exp\left(-\frac{\log n}{16}\right) = o(1)$$

If T denotes the time at which we find the optimal solution:

$$\mathbf{P}[T \leq t_n] \leq \mathbf{P}\left[\sum_{j=1}^{n} X_{j,t_n} \leq \frac{\log n}{4}\right] = o(1)$$

In the end:

$$\mathbf{E}[T] \geq \mathbf{P}[T > t_n]t_n = (1 - o(1))\left(1 - \frac{\log\log n}{\log n}\right)(n^2 - 1)\log n \sim n^2 \log n \quad \square$$

Remark. We are in a very different regime from what happens for $f : \{0,1\}^{n \log_2 n} \to \mathbb{R}_+$. Getting yes/no answers from groups of $\log_2 n$ pieces together somehow results in an additional n factor in the efficiency of $(1+1)$-EA.

We now give an upper bound for Erdos-Renyi method:

Theorem.

$$\mathbf{E}[\mathcal{T}(ER, MM_n)] \leq 2en(\log n + 1)$$

Proof. The probability that the color $c > 0$ at position $1 \leq p \leq n$ was not chosen in any sample x^i such that $f(x^i) = 0$ and $1 \leq i \leq t$ is upper bounded by:

$$\left(1 - \left(1 - \frac{1}{n}\right)^{n-1} \frac{1}{n}\right)^t \leq \exp\left(-t\left(1 - \frac{1}{n}\right)^{n-1} \frac{1}{n}\right) \leq \exp\left(-\frac{t}{en}\right)$$

Let T be the first point in time when all nonzero colors at every position have been chosen in a sample that evaluated to zero. Then by a union bound:

$$\mathbf{P}[T \geq t] \leq n^2 \exp\left(-\frac{t}{en}\right)$$

Let $t_n \stackrel{\text{def}}{=} 2en \log n$. Then:

$$\mathbf{E}[T] \leq \sum_{t=1}^{t_n} \mathbf{P}[T \geq t] + \sum_{t > t_n} \mathbf{P}[T \geq t] \leq t_n + n^2 \frac{e^{-\frac{t_n}{en}}}{1 - e^{-\frac{1}{en}}} = t_n + \frac{1}{1 - e^{\frac{1}{en}}} \leq 2en \log n + 2en$$

$\square$