before re-starting with a new random assignment. However, we will not focus on random restarts in our experiments below because *any* search strategy can be improved (or at the very least, not damaged) by choosing an appropriate cutoff value t [9, 10]. In fact, it is straightforward and well known how to do this optimally (in principle): For a given search strategy and problem, let the random variable T denote the number of flips needed to reach a solution in a single run, and let T_t denote the number of flips needed when using a random restart after every t flips. We then have the straightforward equality [18]

$$\mathrm{E}T_t \;=\; \frac{t}{\mathrm{P}(T \leq t)} - [t - \mathrm{E}(T|T \leq t)] \tag{1}$$

which simply states that $\mathrm{E}T_t$ is t times the expected number of restarts needed to find a solution ($t/\mathrm{P}(T \leq t)$), minus a small correction which accounts for the fact that on the last restart the search will generally find the solution before the maximum cutoff value has been reached. Note that this always offers a potential improvement in principle, since one could choose the optimal cutoff value

$$t^* \;=\; \arg \min_{t: 0 < t \leq \infty} \mathrm{E}T_t$$

which will yield a substantial reduction in expected search time for most natural search distributions [10] and, at the very least, will not increase the expected search time for any distribution (see Appendix A for a detailed discussion). Therefore, we will focus on investigating the single run characteristics of the various variable selection policies, but be sure to report estimates of what the optimum achievable performance would be using random restarts. We report this optimal quantity for every procedure using the empirical distribution of T over several runs (at least 100) to estimate $\mathrm{E}T_{t^*}$.

3 Measuring local search performance

In order to tune the parameters of a search strategy, determine whether a strategic innovation is helpful, or even debug an implementation, it would be useful to be able to measure how well a search is progressing without having to run it to completion on large, difficult problems.

A Effects of random re-starts

It can be shown that imposing an ideal re-start value t^* on the random search time T yields an improvement in expected time for most natural search distributions. To see this, note that for any random variable T we have

$$\mathrm{E}T \;=\; \mathrm{E}(T|T\le t)\;\mathrm{P}(T\le t) \;+\; \mathrm{E}(T|T>t)\;\mathrm{P}(T>t) \tag{2}$$

From (1) we know that the expected search time when using a random restart after t flips is

$$\begin{aligned}
\mathrm{E}T_t &= \frac{t}{\mathrm{P}(T\le t)} - [t - \mathrm{E}(T|T\le t)] \\
&= t\left(\frac{1}{\mathrm{P}(T\le t)} - 1\right) + \mathrm{E}(T|T\le t) \\
&= t\,\frac{\mathrm{P}(T>t)}{\mathrm{P}(T\le t)} + \mathrm{E}(T|T\le t) \qquad (3)
\end{aligned}$$

To determine whether (3) offers an improvement over (2) first note that for an exponential random variable T (i.e., such that $\mathrm{P}(T>x) = e^{-x/\mu}$ for all $x > 0$) we actually have $\mathrm{E}T = \mathrm{E}T_t$ for every cutoff value $t > 0$. This follows from the "memoryless property" of exponential random variables which states that $\mathrm{P}(T>t+x|T>t) = \mathrm{P}(T>x)$ for all $t > 0$ and $x > 0$ [19], and immediately implies that

$$\mathrm{E}(T|T>t) \;=\; t + \mathrm{E}T \tag{4}$$

for all $t > 0$. Thus

$$\begin{aligned}
\mathrm{E}T &= \mathrm{E}(T|T>t)\;\mathrm{P}(T>t) \;+\; \mathrm{E}(T|T\le t)\;\mathrm{P}(T\le t) \\
&= t\;\mathrm{P}(T>t) \;+\; (\mathrm{E}T)\;\mathrm{P}(T>t) \;+\; \mathrm{E}(T|T\le t)\;\mathrm{P}(T\le t) \qquad (5)
\end{aligned}$$

from the memoryless property (4), and therefore for exponential random variables we always have

$$(\mathrm{E}T)\;\mathrm{P}(T\le t) \;=\; t\;\mathrm{P}(T>t) \;+\; \mathrm{E}(T|T\le t)\;\mathrm{P}(T\le t)$$

and hence

$$\begin{aligned}
\mathrm{E}T &= t\frac{\mathrm{P}(T>t)}{\mathrm{P}(T\le t)} \;+\; \mathrm{E}(T|T\le t) \\
&= \mathrm{E}T_t \qquad (6)
\end{aligned}$$

for all $t > 0$, as stated.

So imposing a random restart after a cutoff value t does not affect the expectation of any exponential random variable T. It might therefore appear that random restarts may not offer a useful improvement in general. However, the equality (6) only holds because of the memoryless property (4), and it turns out that this property holds for all $t > 0$ *only* for exponential random variables [19].

By contrast, it is widely believed that the search time distributions for heuristic search algorithms exhibit a "heavy tailed" behavior on difficult constraint satisfaction problems [10]. That is, the tail of the search time distribution $\mathrm{P}(T > x)$ is not exponential, but instead is typically a much slower converging function such as a power law $\mathrm{P}(T > x) = Cx^{-\alpha}$ for $\alpha > 0$, $C > 0$ (where the tails become heavier for smaller α) [10]. A simple example of a power law distribution is a Pareto density $p(x) = \alpha(1+x)^{-\alpha-1}$, $x > 0$, $\alpha > 0$, which defines a random variable T such that $\mathrm{P}(T > x) = (1 + x)^{-\alpha}$. The kth moments of power law distributions are only defined for $k < \alpha$, and in particular their expected values become infinite for $\alpha \leq 1$.

For heavy tailed distributions in general, it is easy to demonstrate that a random re-start strategy will always yield a reduction in expected search time. To see this, first consider the case of a heavy tailed distribution with a finite expected search time; that is, a power law distribution such that $\alpha > 1$. An interesting property of such a distribution is that it is no longer memoryless. In fact, the additional expected search time actually *increases* given that an early solution has not been found:

$$\mathrm{E}(T|T > t) \quad > \quad t + \mathrm{E}T \tag{7}$$

for $t > 0$, which is in direct contrast to (4). From this property it immediately follows that a random re-start strategy yields an improvement in expected solution time, since

$$\begin{aligned} \mathrm{E}T &= \mathrm{E}(T|T > t)\ \mathrm{P}(T > t) \ + \ \mathrm{E}(T|T \leq t)\ \mathrm{P}(T \leq t) \\ &> t\ \mathrm{P}(T > t) \ + \ (\mathrm{E}T)\ \mathrm{P}(T > t) \ + \ \mathrm{E}(T|T \leq t)\ \mathrm{P}(T \leq t) \\ &= \mathrm{E}T_t \end{aligned}$$

For example, for a Pareto distribution with $\alpha > 1$ we have $\mathrm{E}(T|T > f) = t + \mathrm{E}T + t\mathrm{E}T$ for all $t > 0$ and hence immediately obtain $\mathrm{E}T_t < \mathrm{E}T$.

For the more extreme case of a heavy tailed distribution with an *infinite* expected value (i.e., when $\alpha \le 1$) it is obvious that a random re-start strategy significantly improves search time, since $\mathrm{E}T_t$ is clearly finite (3) for any t such that $\mathrm{P}(T \le t) > 0$ by inspection.

Therefore, in general, using a random restart strategy with a cutoff value

$$t^* \;=\; \arg\min_{t:0<t\le\infty} \mathrm{E}T_t$$

(where $\mathrm{E}T_\infty$ is defined to be $\mathrm{E}T$) should improve expected search time, since if (7) is satisfied for any t we will immediately obtain $\mathrm{E}T_{t^*} < \mathrm{E}T$, and otherwise if no such value of t exists we obtain $\mathrm{E}T_{t^*} = \mathrm{E}T$ (and hence cause no harm). Therefore, under any circumstance, employing a random restart strategy after an ideal cutoff value t^* is never a losing strategy.

B Properties of clause re-weighting scheme

Proof of Property 1. First, to show that $\alpha_i F + \beta_i S = 1$, note

$$\begin{aligned}
\alpha_i F + \beta_i S &= \alpha_i F + \frac{1 - F\alpha_i}{S} S \\
&= \alpha_i F + 1 - F\alpha_i \\
&= 1
\end{aligned}$$

Next, to verify that $\alpha_i > 1$, recall that by assumption $\Delta_{x_i} = f_i^+ + s_i^+ - s_i^- < 0 < \delta$, and therefore

$$\begin{aligned}
\alpha_i &= \frac{s_i^- - s_i^+ + S\delta}{F(s_i^- - s_i^+) + Sf_i^+} \\
&> \frac{s_i^- - s_i^+ + S(f_i^+ + s_i^+ - s_i^-)}{F(s_i^- - s_i^+) + Sf_i^+} \\
&= \frac{(1-S)(s_i^- - s_i^+) + Sf_i^+}{F(s_i^- - s_i^+) + Sf_i^+} \\
&= \frac{F(s_i^- - s_i^+) + Sf_i^+}{F(s_i^- - s_i^+) + Sf_i^+} \\
&= 1
\end{aligned}$$