Proof of the Master Theorem

Clément Canonne

COMPx270

Theorem 1 (Master Theorem)**.** Let $T(n)$ be defined by

$$T(n) = \begin{cases} aT\left(\frac{n}{b}\right) + f(n) & \text{if } n \geq n_0 \\ c & \text{if } n \leq n_0 \end{cases} \tag{1}$$

where $a \geq 1, b \geq 2, c \geq 1$ are constants (independent of n), $f(n)$ is a function of n, and n_0 is a positive integer. Then, defining $c^* := \log_b a.$, we have the following:

1. If there exists some positive constant $\varepsilon > 0$ such that $f(n) = O(n^{c^*-\varepsilon})$, then $T(n) = \Theta(n^{c^*})$.
   [Case 1: the recursion dominates]

2. If there exists $k \geq 0$ such that $f(n) = \Theta(n^{c^*}(\log n)^k)$, then $T(n) = \Theta(n^{c^*}\log^{k+1} n)$.
   [Case 2: the recursion and the extra work balance out]

3. If there exists some positive constant $\varepsilon > 0$ such that $f(n) = \Omega(n^{c^*+\varepsilon})$ and [extra "regularity condition"] there is a positive constant $\delta < 1$ such that $af(n/b) \leq \delta f(n)$, then $T(n) = \Theta(f(n))$.
   [Case 3: the extra work dominates]

If $f(n)$ is only given as $O(f(n))$ (e.g., $f(n) = O(n)$, not $3n$ or $\Theta(n)$), then replace all $\Theta(\cdot)$ in the conclusions by $O(\cdot)$. *(If you only have an upper bound on $f(n)$, all conclusions become upper bounds.)*

Note that, for simplicity, we do not explicitly mention floor or ceilings in (1), even though there technically should be: $T\left(\lfloor \frac{n}{b} \rfloor\right)$ or $T\left(\lceil \frac{n}{b} \rceil\right)$. These are annoying, but technically necessary; we ignore them for simplicity.

Proof. We will make two simplifying assumptions (but reasonable ones): first, that $T(n)$ is *non-decreasing in n*: the larger the size of the input, the longer time it takes. The second is that $T(n)$ does not grow *too fast*: $T(n)$ is polynomial in n. The reason is that, first, we shouldn't encounter other things in this class, and second, that we want to be able to say that $T(n)$ and $T(2n)$ differ by at most a constant factor (which is *not* true for, e.g., $T(n) = 2^n$, since then $T(2n) = 2^{2n} = T(n)^2$.

With that in mind, note that if $n \geq 1$ and $m := \lceil \log_b n \rceil$ is the smallest power of b at least n, then by our first assumption

$$T(b^{m-1}) \leq T(n) \leq T(b^m)$$

and the LHS and RHS only differ by a constant factor by the second assumption. What this means is that it suffices to analyse inputs of the form $n = b^m$ (powers of b) to get the answer for general n (not necessarily powers of b). *So let's do that.*

Fix any $m \geq 1$, and observe that then (1) can be rewritten as

$$T(b^m) = aT\left(b^{m-1}\right) + f(b^m)$$

or, equivalently, dividing both sides by a^m,

$$\frac{T(b^m)}{a^m} = \frac{T\left(b^{m-1}\right)}{a^{m-1}} + \frac{f(b^m)}{a^m}$$

Let's define $t_m := \frac{T(b^m)}{a^m}$. Then, the above can be rewritten as

$$u_m - u_{m-1} = \frac{f(b^m)}{a^m}, \qquad \forall m \geq 1 \tag{2}$$

Since the inequality (2) holds for every m, we can sum it for m ranging from 1 to M so get (given the telescoping sum),

$$u_M - u_0 = \sum_{m=1}^{M}(u_m - u_{m-1}) = \sum_{m=1}^{M} \frac{f(b^m)}{a^m}$$

Recalling that $u_M = \frac{T(b^M)}{a^M}$ and that $u_0 > 0$ is "just a constant", what this gives is

$$T(b^M) = a^M \cdot u_0 + a^M \sum_{m=1}^{M} \frac{f(b^m)}{a^m}.$$

Why is this saying? Well, for $n = b^M$, we have $a^M = (b^{\log_b a})^M = (b^M)^{\log_b a} = n^{c^*}$, so this is saying

$$T(n) = \Theta(n^{c^*}) + a^M \sum_{m=1}^{M} \frac{f(b^m)}{a^m}. \tag{3}$$

This is where the 3 cases of the Master Theorem show up. We have 2 terms in the right-hand-side (RHS), and the answer depends on which one of these two terms dominates (is the largest). But that's really hard to figure out! **The three cases are just (sufficient) conditions which allow us to analyse the quantity $\sum_{m=1}^{M} \frac{f(b^m)}{a^m}$ and decide which one of the two terms is bigger.**

1. **Case 1.** Suppose there exists *some* $\varepsilon > 0$ such that $f(n) = O(n^{c^*-\varepsilon})$. Then we have, since $b^{c^*} = b^{\log_b a} = a$,

$$\sum_{m=1}^{M} \frac{f(b^m)}{a^m} = \sum_{m=1}^{M} \frac{O((b^m)^{(c^*-\varepsilon)})}{a^m} = O\left(\sum_{m=1}^{M} \frac{(b^{c^*})^m b^{-\varepsilon m}}{a^m}\right) = O\left(\sum_{m=1}^{M} \frac{a^m b^{-\varepsilon m}}{a^m}\right) = O\left(\sum_{m=1}^{M} (b^{-\varepsilon})^m\right) = O(1)$$

where the last equality follows from the fact that $b^{-\varepsilon} = 1/b^{\varepsilon} \in (0,1)$, so the geometric series converges: $\sum_{m=1}^{M}(b^{-\varepsilon})^m \leq \sum_{m=1}^{\infty}(b^{-\varepsilon})^m < \infty$ is a constant. Recalling (3), since $a^M = n^{c^*}$, we get

$$T(n) = \Theta(n^{c^*}) + n^{c^*} \cdot O(1) = \Theta(n^{c^*}) + O(n^{c^*})$$

so in this case, we have $\boxed{T(n) = \Theta(n^{c^*})}$.

2. **Case 2.** Suppose there exists $k \geq 0$ such that $f(n) = \Theta(n^{c^*}(\log n)^k)$. Then we have $f(b^m) = \Theta(b^{c^* m}(m \log b)^k) = \Theta(a^m m^k)$ (since $(\log b)^k$ is a constant), and so

$$\sum_{m=1}^{M} \frac{f(b^m)}{a^m} = \sum_{m=1}^{M} \frac{\Theta(a^m m^k)}{a^m} = \Theta\left(\sum_{m=1}^{M} m^k\right) = \Theta(M^{k+1})$$

($\sum_{i=1}^{M} m^k = \Theta(M^{k+1})$, sums behave like integrals: $\int_0^x t^k dt = x^{k+1}/(k+1)$). Recalling (3), since $a^M = n^{c^*}$ and $M = \log_b n$, we get

$$T(n) = \Theta(n^{c^*}) + n^{c^*} \cdot \Theta((\log_b n)^{k+1}) = \Theta(n^{c^*}) + \Theta(n^{c^*}(\log n)^{k+1})$$

so in this case, we have $\boxed{T(n) = \Theta(n^{c^*}(\log n)^{k+1})}$.

3. **Case 3.** Suppose now there exists some positive constant $\varepsilon > 0$ such that $f(n) = \Omega(n^{c^*+\varepsilon})$ (we'll see soon where we also need the "regularity condition"). Then

$$\sum_{m=1}^{M} \frac{f(b^m)}{a^m} = \sum_{m=1}^{M} \frac{\Omega((b^m)^{(c^*+\varepsilon)})}{a^m} = \Omega\left(\sum_{m=1}^{M} \frac{a^m b^{\varepsilon m}}{a^m}\right) = \Omega\left(\sum_{m=1}^{M} (b^{\varepsilon})^m\right) = \Omega((b^{\varepsilon})^M) = \Omega(n^{\varepsilon})$$

the second-to-last equality since now the geometric series is divergent, so the sum is dominated by its last term. This means that the *second term* of (3) dominates, as we have

$$T(n) = \Theta(n^{c^*}) + n^{c^*} \cdot \Omega(n^{\varepsilon}) = \Theta(n^{c^*}) + \Omega(n^{c^*+\varepsilon})$$

Now, that's good, the second term dominates…but what is it equal to? This is where the regularity condition, that there is some $\delta \in [0,1)$ such that $af(n/b) \leq \delta f(n)$ for all n, will be useful. Indeed, we can write that second term as

$$a^M \sum_{m=1}^{M} \frac{f(b^m)}{a^m} = \sum_{m=1}^{M} a^{M-m} f(b^m) = \sum_{j=0}^{M-1} a^j f(b^{M-j})$$

and the first term ($j = 0$) is exactly $f(b^M) = f(n)$. If we could say that the whole sum is basically dominated by this first term, the whole thing would be $\Theta(f(n))$, and we'd be done.

But we can! By using the regularity condition to get something related to a geometric series, *again*. Indeed:

$$a^j f(b^{M-j}) = a^{j-1} \cdot af(b^{M-j+1}/b) \leq a^{j-1} \cdot \delta f(b^{M-j+1}) \leq \cdots \leq \delta^j f(b^M)$$

repeatedly using this regularity condition. So we can write:

$$\begin{aligned}
\sum_{j=0}^{M-1} a^j f(b^{M-j}) &= f(b^M) + \sum_{j=1}^{M-1} a^j f(b^{M-j}) \\
&\leq f(b^M) + \sum_{j=1}^{M-1} \delta^j f(b^M) \leq f(b^M) + f(b^M) \sum_{j=1}^{\infty} \delta^j \\
&= f(b^M) + f(b^M) \cdot O(1) = \Theta(f(b^M))
\end{aligned}$$

since $\sum_{j=1}^{\infty} \delta^j < \infty$ is just a constant (converging geometric series). **So** the second term in (3) not only dominates, it's also $\Theta(f(b^M))$ (and $b^M = n$)! And that means we're done: $\boxed{T(n) = \Theta(f(n))}$.

□

Note that these 3 cases are not actually comprehensive! There are other possibilities not covered by the theorem…the key is that these 3 cases provide (reasonably useful) conditions under which the second term of (3) can be analysed.