Sorting vs. Braids vs. The Substitution Lemma vs. $\lambda\mathtt{xc}$

Knowledge of the book 'Term Rewriting Systems', Terese, CUP, 2003 is assumed.

Sorting Sorting is the process of stepwise transforming a *source* order into a *target* order on the same objects. It is *swapping*-based if in each step the order changes for exactly one pair of objects. In this note, the rewrite properties of swapping-based sorting are studied. To that end, swapping-based sorting is formalised as a rewrite system. Its objects, called multi-swaps, will be pairs of relations which can be thought of as the 'current order' and the 'remaining inversions'.

Definition 1. *A* state *is a total irreflexive transitive relation on the set of positions of a finite string. A* multi-swap *is a pair consisting of a state $<$ and a transitive scopic subrelation T of it, where T is* scopic *if $p < o < q$ and $p\ T\ q$ imply $p\ T\ o$ or $o\ T\ q$. A* swap *is a multi-swap $(<, \widehat{pq})$, where $\widehat{pq}$ is given by $p\ \widehat{pq}\ q$, implicitly assuming p* adjacent *to q, i.e. there is no o with $p < o < q$.*

The two coherence conditions on the 'remaining inversions' T in the definition are due to Melliès (RTA 2002). They can be thought of as discrete axiomatisations of continuous phenomena:[1] Imagine three guys p, o, and q standing in a row, o being the man in the middle, which start walking column-wise. Then scopic captures that for p and q to cross paths, either p or q has to cross paths with o. Dually, transitive captures that if both p and q cross paths (exactly once!) with o, then p and q have to cross paths. Without these conditions the 'remaining inversions' could be incoherent. Imagine for example, someone requesting to sort a list $[p, o, q]$ saying that both pairs p, o and o, q are already in the correct order, but p and q are not. Or, for another example, that p, q are in the correct order, but both p, o and o, q should be swapped? As a first sanity check, bijections between orders, the big-step semantics of sorting so to speak, are modeled as multi-swaps.

Lemma 2. *For any states $<$, $\ll$ on the same set of objects $(<, < - \ll)$ is a multi-swap.*

Proof. By T the relation $< - \ll$, which captures the idea that the 'remaining inversions' are exactly those pairs in $<$ which are not (yet) in the correct order w.r.t. $\ll$, is abbreviated. It must be verifyied that $(<, T)$ is a multi-swap. Clearly T is a subrelation of $<$. To show T is transitive, suppose $p\ T\ o\ T\ q$. Then $p < o < q$, and $q \ll o \ll p$ by totality of $\ll$. Thus $p < q$ and $q \ll p$ by transitivity twice, and one concludes $p\ T\ q$ from totality of $\ll$. To show T is scopic, suppose $p < o < q$ and $p\ T\ q$. Then $p < q$ and $q \ll p$. By totality of $\ll$, either $p \ll o$ or $o \ll p$. In case the former, $q \ll o$ by transitivity, hence $o\ T\ q$. In case the latter, $p\ T\ o$. □

Below this modeling is shown correct. The relation $< - \ll$ can be visualised by drawing twice, below each other, the set of objects on a line, ordered from left to right according to $<$ and $\ll$, respectively. Connecting each pair of copies of the same object by a straight line, the relation T consists of all pairs in $<$ whose paths cross. The top-down symmetry in the visualisation immediately suggests the equality $(< - \ll) = (\gg - >)$, which indeed holds as one easily verifies by calculation. For any state $<$, $(<, \emptyset)$ and $(<, <)$ are respectively its *empty* and *full* multi-swap (note $<$ is scopic). The empty multi-swap corresponds to a bijection from $<$ to itself, and the full multi-swap to a bijection from $<$ to $>$. As a second sanity check, it is proven that any coherent collection of 'remaining inversions' must contain a pair of swappable, i.e. adjacent, objects.

Lemma 3. *Every non-empty multi-swap contains at least one swap.*

Proof. Let $(<, T)$ be a non-empty multi-swap. Define the *diameter* of a pair $p < q$ to be the length n of the longest sequence $o_1 < \ldots < o_n$ with $p = o_1$ and $o_n = q$. The diameter is well-defined by finiteness, transitivity and irreflexivity of $<$. The result follows by well-founded induction on the diameter of pairs in T, using that the latter is scopic. □

[1] One may think of the Jordan Curve Theorem.

Viewing a multi-swap as consisting of a string (the state) to be sorted accoding to a set of rewrite rules of the form $axb \to bxa$ (the remaining inversions), the lemma expresses that always some rule of the form $ab \to ba$ is applicable. Next, the effect of a swap is defined. On the current order it does what it says; it swaps the order of the objects involved. W.r.t. the remaining inversions it is simply removed; it has been done.

Definition 4. *The* residual *of* $(<, T)$ after $(<, \widehat{pq})$ *is* $((< - \widehat{pq}) \cup \widehat{qp}, T - \widehat{pq})$.

Observe that if T is contained in S for multi-swaps $(<, T)$ and $(<, S)$, then the residual of the former multi-swap after some swap is contained in the residual of the latter after the same swap. Next, the soundness of the notion of coherence w.r.t. swapping is checked: the remaining inversions should still be coherent after performing a swap.

Lemma 5. *The residual of a multi-swap after a swap contained in it, is a multi-swap.*

Proof. Let $(\ll, S)$ be the residual of the multi-swap $(<, T)$ after the swap $(<, \widehat{pq})$. Totality and irreflexivity of $\ll$ follow easily from the same for $<$. To show transitivity of $\ll$, let $p' \ll o \ll q'$. Then $p' < o < q'$, hence $p' < q'$ by transitivity of $<$ and thus $p' \ll q'$ unless $p = p'$ and $q = q'$. However, the latter cannot both hold as then o would contradict p and q being adjacent. Similarly transitivity of S follows from transitivity of T. Finally, the only interesting case (up to symmetry) in showing that S is scopic is if $q \ll p \ll o$ and $q \; S \; o$. Then $p \; T \; q \; T \; o$, so $p \; T \; o$ and $p \; S \; o$. ☐

Soundness of swaps having been verified sorting is modeled as rewriting.

Definition 6. *Let* $\hookrightarrow$ *be the* swapping *ARS having multi-swaps for objects and, for any multi-swap and swap contained in it, a step from the former to its residual after the latter.*

As one would hope and expect, sorting by swapping always terminates in a unique result.

Theorem 7. $\hookrightarrow$ *is complete.*

Proof. By Newman's Lemma, termination and local confluence suffice to conclude completeness. Termination of $\hookrightarrow$ follows noting that the number of pairs in the second component of a multi-swap decreases in each step. To show local confluence, consider steps induced by distinct swaps $(<, \widehat{pq})$ and $(<, \widehat{uv})$. If $q = u$ then first swap $\widehat{pv}$ on both reducts, and then $\widehat{qv}$ on the former and $\widehat{pq}$ on the latter, to reach a common reduct. Otherwise, swap $\widehat{uv}$ respectively $\widehat{pq}$. ☐

Note that if $\hookrightarrow$ would have been modeled as a string rewriting system in the way suggested above, checking local confluence reduces by the Critical Pair Lemma to checking joinability of the critical pair corresponding to $q = u$ in the proof of the theorem. Finally, correctness of the earlier modeling of bijections can be both expressed and proven.

Lemma 8. *For any states* $<$ *and* $\ll$, $(\ll, \emptyset)$ *is the normal form of the multi-swap* $(<, < - \ll)$.

Proof. By $\ll = (< - (< - \ll)) \cup (< - \ll)^{-1}$, which just expresses that performing the inversions necessary to go from $<$ to $\ll$ indeed yields $\ll$ when applied to $<$. ☐

That is, sorting w.r.t. some order results in a sorted result. One may observe that the swapping ARS $\hookrightarrow$ is even balanced weak Church–Rosser (BWCR) in the sense of Toyama (LICS 1991), since any peak $\hookleftarrow ; \hookrightarrow$ can be joined in a balanced way by either zero (for the same step), one ($\hookrightarrow ; \hookleftarrow$ for non-overlap), or two ($\hookrightarrow ; \hookrightarrow ; \hookleftarrow ; \hookleftarrow$ for overlap) steps. Toyama's results for BWCR ARSs not only imply that to prove termination of $\hookrightarrow$ it would have sufficed to prove normalisation, i.e. to prove that some swapping *strategy* terminates, but more interestingly that all reductions to normal form have the same length, i.e. that all swapping-based sorting algorithms have the same (quadratic) complexity; a full multi-swap on n objects always takes $n(n-1)/2$ steps. In the following final lemma of this section, it is shown how from consecutive multi-swaps their composition can be constructed.

Lemma 9. *If T, S are consecutive multi-swaps from state $<$ to $\ll$ to $\lll$, then $(T - S^{-1}) \cup (S - T^{-1})$ is one from $<$ to $\lll$.*

Proof. The idea of the definition is that all inversions are taken which are done in either multi-swap minus the 'double' ones. Correctness is calculated as follows.

$$
\begin{aligned}
& (T - S^{-1}) \cup (S - T^{-1}) \\
= \quad & ((< - \ll) - (\ll - \lll)^{-1}) \cup ((\ll - \lll) - (< - \ll)^{-1}) \\
= \quad & ((< - \ll) - (\lll - \ll)) \cup ((\ggg - \gg) - (> - \gg)) \\
= \quad & ((< - \lll) - (\ll - \lll)) \cup ((\ggg - >) - (\gg - >)) \\
= \quad & ((< - \lll) - (\ll - \lll)) \cup ((< - \lll) - (< - \ll)) \\
= \quad & (< - \lll) - ((\ll - \lll) \cap (< - \ll)) \\
= \quad & < - \lll
\end{aligned}
$$

using the equality suggested above and some relation algebra, in particular the cube axiom of residual systems for the third, and disjointness of T and S for the last equality. $\square$

By disjointness of T and S, the lemma, and the idea expressed in its proof, it follows that the parity of the sum of the cardinalities of T and S is the same as that of the direct multi-swap. This is the ordered version of the classical fact that the parity of all decompositions of a given permutation into transpositions is always the same, even or odd; here it is equal to the parity of (the cardinality of) the multi-swap between the two states.

Braids Braids arise by splitting the notion of swapping into *over crossing* and *under crossing*. That is, it now matters whether the one object should be crossed over or under the other. Suppose first a would have been crossed over b in $[a, b]$ resulting in $[b, a_b]$, where the subscript is *ad hoc* notation to indicate a crossed *over* b. Then for braids, continuing with crossing b under a results simply in $[a, b]$, whereas crossing b over a results in $[a_b, b_a]$, which are different results. One can try it with two strands; over crossing twice (a *twist*) is topologically different from first over and then under crossing; the latter is just the identity. In this note, the interest will be in situations where only one form of crossing, say over crossing, is allowed, i.e. in the so-called *braid* semi-group. Replacing in the previous section everywhere *swap* by (*over*) *cross* its results can be lifted without effort, in particular the crossing ARS $\hookrightarrow$ is complete. But more can be said.

Theorem 10. *The ARS $\hookrightarrow$ is* topologically *confluent, i.e. any pair of co-initial reductions can be completed into a confluence diagram having topologically equivalent sides.*

Proof. By Lemma 7, noting that both sides of each of the three local confluence diagrams employed in its proof are topologically equivalent. In fact, the two non-trivial local confluence diagrams *are* the usual equations defining braid equivalence; the case that $q = u$ corresponds to Reidemeister's *slide* move for knots, and to the Yang-Baxter equation in physics (yes, there are many models in physics). $\square$

Whereas two consecutive sorts (multi-swaps) can always be combined into a single one (the composition of two bijections is a bijection again) as shown in the previous section, this is not the case for multi-crossings as witnessed by the example above; a twist cannot be represented as a single multi-crossing simply because it would require crossing the same strands twice, something which is not possible in a multi-crossing. This leads to the following definition of braids as sequences of multi-crossings, i.e. as 'repeated sorting'.

Definition 11. *The* braid *ARS has states as objects and, for any $\hookrightarrow$-reduction sequence ending in the empty multi-crossing, a step from the first component of the source to that of the target.*

Theorem 12. *The braid ARS is topologically confluent.*

Proof. It suffices to show the braid ARS has the Triangle Property. For any state, let its *full* state be the (unique) target of the braid step induced by its full multi-crossing. Now consider an arbitrary braid step induced by some multi-crossing on the state. Since the multi-crossing is contained in the full multi-crossing, the sequence of crossings corresponding to the former can also be applied to the full multi-crossing, as observed above. By completeness of $\hookrightarrow$, the target of that reduction sequence can be further reduced to the full state. □

The Substitution Lemma The results in the previous sections are mild reformulations of known results. In this section the methods developed in the previous sections to show topological confluence of braids are extended to show confluence of a rewrite system derived from the so-called Substitution Lemma of the λ-calculus:

$$M[x{:=}N][y{:=}P] = M[y{:=}P][x{:=}N[y{:=}P]]$$

which arises from the critical pair between the two β-reduction steps which are possible from the term $(\lambda y.(\lambda x.M)N)P$. In explicit substitution calculi the substitution operation is reified into an explicit substitution operator, transforming the equality into an equation having syntactically distinct left- and right-hand sides. Thus, in order to regain confluence, the equation should be 'completed'. The system studied here arises by choosing to orient the equation from left to right, i.e. as

$$M[x{:=}N][y{:=}P] \to M[y{:=}P][x{:=}N[y{:=}P]]$$

The question is whether this single rule is confluent. Since the λ-calculus syntax is rather unwieldy here, the confluence problem is solved for the following simpler syntax, still capturing the essential difficulties.

Definition 13. *Let $\mathcal{S}$ be the* substitution *TRS having a unary operator* $[\_]$ *and an associative binary operator denoted by juxtaposition, and the single rule $\varsigma : [x][y] \to [y][x[y]]$.*

The proof of confluence of this substitution TRS will follow the same structure as the proof of topological confluence of braids in the previous section. The underlying idea is that both crossing two strands and interchanging two substitutions can be thought of as 'commutation with history', and the latter is an extension of the former because it allows nested substitutions, resulting in 'rewriting history' so to speak. The notion of residual to be defined below, will need the preliminary concept of the descendant relation ς_p for an $\mathcal{S}$-step at position p. It is induced in the standard way (see Terese) for TRSs, by the descendant relation for the rule ς indicated by its labeling $[x]^i[y]^{i+1} \to [y]^{i+1}[x[y]^{i+1}]^i$. Note ς_p^{-1} is functional and surjective.

Definition 14. *Let $<$ be the relation on positions given by $piq < pj$, for all p, q, and $i < j$, and let $<_t$ be its restriction to the positions of term t. A* multi-swap *is a pair consisting of a term t and a transitive scopic left-convex subrelation T of $<_t$, where T is* left-convex *if $piq\ T\ pj$ implies T relates (each position on) the path between pi and piq (inclusive) to pj. A* swap *is a multi-swap $(t, \widetilde{pi})$ where $\widetilde{pi}$ is given by $pi\ \widetilde{pi}\ p(i+1)$.*

For any term t, $(t, \emptyset)$ and $(t, <_t)$ are respectively its *empty* and *full* multi-swap (note $<$ is trivially irreflexive, transitive, and left-convex, and $<_t$ is moreover finite and scopic, for any term t).

Lemma 15. *Every non-empty multi-swap contains at least one swap.*

Proof. Let (t, T) be a non-empty multi-swap. By non-emptiness, by $T \subseteq <_t \subseteq <$, and by definition of $<$, $piq\ T\ pj$ for some p, q, and $i < j$. Therefore, by T being left-convex, also $pi\ T\ pj$. Since $pi <_t o <_t pj$ implies by definition of $<$ that all three positions are at the same *level*, i.e. $o = pk$ for some k, one may conclude as in Lemma 3, noting the positions are adjacent iff $j = i + 1$. □

Definition 16. *The* residual *of (t, T)* after *$(t, \widetilde{oi})$ is $(s, (\varsigma_{oi}^{-1} ; T ; \varsigma_{oi}) \cap (<_s - \widetilde{oi}))$, if $t \to_{\mathcal{S},oi} s$.*

One can think of this definition of residual of T in linear algebraic terms as the ς_{oi}-conjugate of T w.r.t. spaces (of positions) t and s.

Lemma 17. *The residual of a multi-swap after a swap contained in it, is a multi-swap.*

Proof. Let (s, S) be the residual of the multi-swap (t, T) after $(t, \widetilde{oi})$. By definition of S as an intersection with (a subrelation of) $<_s$, the former is a subrelation of the latter.

To show transitivity of S, suppose $p\ (S\,;S)\ q$. This implies $p\ (\varsigma_{oi}^{-1}\,; T\,; \varsigma_{oi})\ q$, as $\varsigma_{oi}\,; \varsigma_{oi}^{-1}$ is the identity relation and T is transitive by assumption, and also $p <_s q$ by transitivity of $<_s$. Finally, if $p\ \widetilde{oi}\ q$ were to hold, then $p = oi$ and $q = o(i+1)$, entailing $o(i+1)\ T\ oi$ contradicting $T \subseteq <_t$.

To show S is left-convex, suppose $pjq\ S\ pk$, so $j < k$ must hold since $S \subseteq <_s$. It suffices to show that if some prefix v of pjq is related by S to pk, then so is its predecessor u if it is a suffix of pj. To that end, let ς_{oi} relate p', v', and u' to pk, v and u, respectively. Then, by definition of S it holds $v'\ T\ p'$, and by the form of ς, u' is seen to be the predecessor of v' unless $u = o(i+1)$ and $v = o(i+1)i'$ with i' the maximal successor (the position of the nested $[y]$). In case the latter, $u' = oi\ T\ o(i+1) = v'\ T\ p'$, hence by transitivity of T, $u'\ T\ p'$ so $u\ S\ pk$. In case the former, let $q'\ \varsigma_{oi}\ pj$. Then, as before, q' is a prefix of u' unless $pj = o(i+1)$ and $u = o(i+1)i'$. In case the latter, $u' = o(i+1) = pj < pk = p'$, hence as u' is a prefix of v', $u'\ T\ p'$ by left-convexity of T, so $u\ S\ pk$. In case the former, q' and p' are on the same level unless $pj = o(i+1)j'$ for some $j' < i'$ and $pk = o(i+1)i'$. In case the latter, $q' = o(i+1)$ and since pi is a prefix of $pij' = q'$ which is a prefix of u', one has $u'\ T\ p'$ by left-convexity of T, so $u\ S\ pk$. In case the former, q' is to the left of p' unless $pj = oi$ and $pk = o(i+1)$. In case the latter, $p' = oi$ and $o(i+1) = q'$ is a prefix of v' which is a prefix of u' contradicting $p' <_t u'$ hence $p'\ T\ u'$. In case the former, $u'\ T\ p'$ by left-convexity of T, so $u\ S\ pk$.

To show S is scopic, suppose $p <_s o' <_s q$ and $p\ S\ q$. Let ς_{oi} relate p', o'' and q' to p, o' and q, respectively, so $p'\ T\ q'$ by definition of $p\ S\ q$. If $p' <_t o'' <_t q'$ either $p'\ T\ o''$ or $o''\ T\ q'$ since T is scopic. If $p'\ T\ o''$, one concludes $p\ S\ o'$ from $p\ (\varsigma_{oi}^{-1}\,; T\,; \varsigma_{oi})\ o'$ and $p <_s o'$, since $p\ \widetilde{oi}\ o'$ would entail $o'' = oi$ and $p' = o(i+1)$ contradicting $p' <_t o''$. The case $o''\ T\ q'$ follows symmetrically. It remains to consider situations in which either $p' <_t o''$ or $o'' <_t q'$ does not hold. If $p' <_t o''$ does not hold, then $o'' = oi$, $o(i+1)$ is a prefix of p', and $o(i+1) = o' < q = q'$. Hence $o(i+1)\ T\ q'$ follows from left-convexity of T for $p'\ T\ q'$, using that the greatest common prefix of p and q is a prefix of o by $o' <_s q$, hence the same as the greatest common prefix of p' and q', which is therefore a prefix of $o(i+1)$. Since also $oi\ T\ o(i+1)$, transitivity of T yields $o''\ T\ q'$, so $o'\ S\ q$. If $o'' <_t q'$ does not hold, then $q' = oi$, and $o(i+1)$ is a prefix of o''. Note in fact it must hold $o'' = o(i+1)$, since otherwise $o(i+1)$ would be a prefix of p', contradicting $p'\ T\ q'$. Thus $p'\ T\ q' = oi\ T\ o(i+1) = o''$, hence $p'\ T\ o''$ by transitivity of T, so $p\ S\ o'$. $\square$

Lemma 18. $\hookrightarrow$ *is complete, for* $\hookrightarrow$ *defined as in Definition 6.*

Proof. By Newman's Lemma, termination and local confluence suffice to conclude completeness. Termination of $\hookrightarrow$ follows by measuring a term t by the multiset of lengths of all paths of maximal diameter in $<_t$. By the Critical Pair Lemma, to show local confluence it suffices to consider the critical pair arising from $([x][y][z], \{(1,2),(2,3),(1,3)\})$. It is confluent as can be seen by enriching both $[x]\underline{[y][z]} \hookrightarrow \underline{[x][z]}[y[z]] \hookrightarrow [z]\underline{[x[z]][y[z]]} \hookrightarrow [z][y[z]][x[z][y[z]]]$ and $\underline{[x][y]}[z] \hookrightarrow [y]\underline{[x[y]][z]} \hookrightarrow \underline{[y][z]}[x[y][z]] \hookrightarrow [z][y[z]][x\underline{[y][z]}] \hookrightarrow [z][y[z]][x[z][y[z]]]$ with relations. $\square$

Thus, the main result is obtained.

Theorem 19. $\mathcal{S}$ *is confluent.*

Proof. Completely analogous to the proof of Theorem 12. $\square$

The TRS $\mathcal{S}$ is easily seen to be isomorphic to *combinatorless* combinatory logic, where terms consist only of applications and variables, having the *S-less* S-rule $wxy \to wy(xy)$ as only rule. By the theorem, combinatorless combinatory logic is confluent.

$\lambda\mathtt{xc}$ Extending Bloo and Rose's $\lambda\mathtt{x}$-calculus with the substitution lemma rule of the previous section, yields the so-called $\lambda\mathtt{xc}$-calculus.

Definition 20. *The $\lambda\mathtt{xc}$-calculus is given by the rules*

$$\begin{array}{rcl}
(\lambda y.X)Y & \to & X[y{:=}Y] \\
x[y{:=}Y] & \to & x \\
y[y{:=}Y] & \to & Y \\
(X_1X_2)[y{:=}Y] & \to & (X_1[y{:=}Y])(X_2[y{:=}Y]) \\
(\lambda x.X)[y{:=}Y] & \to & \lambda x.X[y{:=}Y] \\
X[y{:=}Y][z{:=}Z] & \to & X[z{:=}Z][y{:=}Y[z{:=}Z]]
\end{array}$$

It will turn out handy to split he set of rules as follows. The first rule is the $\mathtt{Beta}$-rule, the next four are the $\mathtt{x}$-rules, and the final rule is the $\mathtt{c}$-rule. Any $\lambda\mathtt{xc}$-term can be reduced to *substitution* normal form, i.e. $\mathtt{x}$-normal form, yielding a term without explicit substitutions, i.e. an ordinary λ-term. This is the basis for Hardin's so-called *interpretation* method to reduce confluence of λ-calculi with explicit substitutions to (the well-known) confluence of the ordinary λ-calculus, which also works for $\lambda\mathtt{x}$. At the same time, this shows that the $\mathtt{c}$-rule is equationally superfluous and thus confluence of the $\lambda\mathtt{xc}$-calculus follows from that of the $\lambda\mathtt{x}$-calculus. However, it might be advantageous to have $\mathtt{c}$ as an optimisation rule. For one, whereas in the $\lambda\mathtt{x}$-calculus it may require exponentially many steps to make a $\mathtt{Beta}$-redex which is present in the substitution normal form of a term t 'visible', in the $\lambda\mathtt{xc}$-calculus that requires at most a linear number of steps. For another, completing the critical pair between the $\mathtt{Beta}$-rule and the $\mathtt{x}$-rules:

$$X[y{:=}Y][z{:=}Z] \leftarrow ((\lambda y.X)Y)[z{:=}Z] \to ((\lambda y.X)[z{:=}Z])(Y[z{:=}Z])$$

may require, in $\lambda\mathtt{x}$, performing all the substitutions which occur in (the term substituted for) X, which may involve exponentially many steps, but with the help of the $\mathtt{c}$-rule it can be completed in just[2] three steps:

$$X[y{:=}Y][z{:=}Z] \to X[z{:=}Z][y{:=}Y[z{:=}Z]] \leftarrow (\lambda y.X[z{:=}Z])(Y[z{:=}Z]) \leftarrow ((\lambda y.X)[z{:=}Z])(Y[z{:=}Z])$$

As a third:

Lemma 21. *$t[y{:=}s] \twoheadrightarrow_{\mathtt{xc}} t$, if y does not occur in t, in a number of steps linear in t.*

Proof. By induction on the size of t distinguishing cases on its shape. If it is of shape $t_1[x{:=}t_2]$, then $t \to_{\mathtt{c}} t_1[y{:=}s][x{:=}t_2[y{:=}s]]$ and one concludes by applying the IH to both $t_i[y{:=}s]$. Otherwise, one of the $\mathtt{x}$ rules (but not the second) is applicable at the head, and one concludes by the IH. ☐

Again, due to the absence of $\mathtt{c}$, this fails in the $\lambda\mathtt{x}$-calculus where both sides are just convertible and that only in an exponential number of steps (to see the latter, consider a term t of shape $(x_1x_1)[x_1{:=}(x_2x_2)][x_2{:=}(x_3x_3)]\ldots[x_n{:=}z]$). The lemma justifies replacing the first $\mathtt{x}$-rule by the so-called *garbage collection* rule $X[y{:=}Y] \to X$. To conclude this note, a direct proof of confluence of the resulting calculus, which we will still refer to as $\lambda\mathtt{xc}$, is given, i.e. without relying on confluence of the ordinary λ-calculus [3]

Lemma 22. *$\lambda\mathtt{xc}$ is confluent.*

Proof. Let $\mathrel{\circ\!\!\!\!\rightarrow}_{\mathtt{Beta}}$ denote the contraction of any number of $\mathtt{Beta}$-redexes, let $\to_{\mathtt{x}}$ denote contracting an arbitrary (possibly garbage collecting) $\mathtt{x}$-redex, and let $\mathrel{\circ\!\!\!\!\rightarrow}_{\mathtt{c}}$ denote the contraction of a multi-swap (in the sense of the previous section) of $\mathtt{c}$-redexes. To show confluence of $\lambda\mathtt{xc}$, it then suffices to prove confluence of these relations, since $\to_{\lambda\mathtt{xc}} \subseteq \mathrel{\circ\!\!\!\!\rightarrow}_{\mathtt{Beta}} \cup \to_{\mathtt{x}} \cup \mathrel{\circ\!\!\!\!\rightarrow}_{\mathtt{c}} \subseteq \twoheadrightarrow_{\lambda\mathtt{xc}}$. In the rest of the proof, the three types of rules are referred to simply as $\mathtt{Beta}$, $\mathtt{x}$, and $\mathtt{c}$. It suffices to show that they are decreasing w.r.t. the well-founded order which orders $\mathtt{Beta}$ above $\mathtt{c}$ above $\mathtt{x}$, and orders $\mathtt{x}$ steps according to the size of their source. Distinguish cases on the types of the rules in a peak.

[2]It was designed for that purpose!

[3]The same method yields a direct proof of confluence of the $\lambda\mathtt{x}$-calculus, without relying on interpretation.

- If both are $\mathtt{Beta}$, then the result follows from single-step $\mathtt{Beta}$ being *linear orthogonal*, i.e. a common reduct is found in either zero or one $\mathtt{Beta}$-steps on both sides.

- If $s \leftarrow_{\mathtt{x}} t \mathrel{\circ\!\!\!\rightarrow}_{\mathtt{Beta}} r$, then $s \rightarrow_{\mathtt{x}} ; \mathrel{\circ\!\!\!\rightarrow}_{\mathtt{Beta}} q$ and $r \mathrel{\circ\!\!\!\rightarrow}_{\mathtt{c}} q$ in case the $\mathtt{x}$-step overlaps one of the $\mathtt{Beta}$-redexes, and $s \mathrel{\circ\!\!\!\rightarrow}_{\mathtt{Beta}} q \twoheadleftarrow_{\mathtt{x}} r$ otherwise.

- $\mathtt{Beta}$ is orthogonal to $\mathtt{c}$, and they commute in a single step on either side.

- If both are $\mathtt{x}$, a common reduct is found in at most two further $\mathtt{x}$-steps having smaller sources.

- If $s \leftarrow_{\mathtt{x}} t \mathrel{\circ\!\!\!\rightarrow}_{\mathtt{c}} r$, then $s \twoheadrightarrow_{\mathtt{x}} ; \mathrel{\circ\!\!\!\rightarrow}_{\mathtt{c}} q \twoheadleftarrow_{\mathtt{x}} r$ in case the $\mathtt{x}$-step overlaps one of the $\mathtt{c}$-redexes (needing a number of garbage collection steps). Otherwise the steps simply commute ($\mathtt{c}$ may duplicate $\mathtt{x}$, but not *vice versa*).

- If both rules are $\mathtt{c}$, then a common reduct is found in at most one further $\mathtt{c}$-step as shown in the previous section. □

Note that a common reduct is found in an amount of work which is linear in the diverging steps, measuring each step by the number of *single* steps performed by it. This is not that good, but still better than the interpretation method.

Of course, a drawback of the $\lambda\mathtt{xc}$-calculus is that it is inherently non-terminating due to the $\mathtt{c}$-rule, even on terms which are terminating in the ordinary λ-calculus w.r.t. β-reduction. But the idea is that the $\mathtt{c}$-rule could be seen as an optimization rule, and the other rules as computation rules; the computation rules are known to preserve the good properties for $\lambda\mathtt{x}$. It remains to show the same holds for the optimization rule. For instance, it would be interesting to know whether it preserves acyclicity (or even stronger, whether $\lambda\mathtt{xc}$ preserves acyclicity).