1 Finiteness of Development

We prove the finiteness of developments for the untyped lambda calculus. We begin by definiting the notion of a residual using the underlined rewrite system.

Definition. Define $\underline{\Lambda}$ inductively as the smallest set containing the set of variables and closed under

- MN
- $\lambda x . M$
- $\underline{(\lambda x . M)N}$

We define a notion of reduction on $\underline{\Lambda}$ called $\underline{\beta}$ by

$$\underline{(\lambda x . M)N} \to_{\underline{\beta}} M[x := N]$$

and make the terms behave as a congruence. We write $\twoheadrightarrow_{\underline{\beta}}$ for the transitive, reflexive closure.

Remark. The funamental observation is that no new unlines are placed when doing a $\underline{\beta}$ reduction, so although redexes can be transformed and copied, new ones cannot be created.

Definition 1. If $\mathscr{F}$ is a set of redexes in a term $M \in \Lambda$ we write $M^{\mathscr{F}}$ for the $\underline{\Lambda}$ term which underlines all the redexes in $\mathscr{F}$.

If $M \in \underline{\Lambda}$ we write $|M|$ for the term in Λ achieved from erasing all the underlines.

If $M \twoheadrightarrow_{\beta} N$ by a reduction path σ and Δ is a redex in M then we define the set of residuals of Δ with respect to σ, which we write Δ/σ, as the set of underlined term in N' when doing the corresponding reduction $\underline{\sigma}$ in $M^{\{\Delta\}} \twoheadrightarrow_{\underline{\beta}} N'$

We say $M \twoheadrightarrow N$ is a development if there is a set of redexes $\mathscr{F}$ from M such that $M^{\mathscr{F}} \twoheadrightarrow_{\underline{\beta}} N'$ and $|N'| = N$. Another way to say this is a development is one in which only residuals of redexes in M are contracted.

Theorem 1 (Finiteness of Developments)**.** *All developments are finite.*

The proof of this amounts to showing that $\underline{\beta}$ is strongly normalizing. We present three proofs.

1.1 Weights

As far as I know this proof is due to Klop and Barendregt and probably some others I'm missing. The ideas are taken from Barendregt's 1980 book.

Definition 2. Given a term $M \in \underline{\Lambda}$ we say the term is weighted if it has associated with it a function from variable instances to the natural numbers. We normally write these numbers as supserscripts on the variables, or by the function wt (although, be careful, because the weight is to all variable instances, so different x's get different weights. Such notation will only be used when it is not confusing what instance we are talking about).

If M is weighted and $M \to_{\underline{\beta}} N$ then we get an associated weighting of N which doesn't require much description; just keep the superscripts around when copying variables.

We can assign a weight all weighted terms M by the summing the weights of all the variables in M. With this we weigth not only weighted terms but also their subterms.

We say that a weighting of M is decreasing if for every redex $\Delta = (\lambda x . P)Q$ in M, for every variable instance of x in P we have that $\mathrm{wt}(x) < \mathrm{wt}(Q)$

Proposition. *Every term can be given a decreasing weight.*

Proof. Beginning from the right, give variables increasing powers of 2 as weights. Then for any redex $(\lambda x . P)Q$ any variable x in P will be weighted 2^i and any variable to the right of x will have weight 2^j for $j < i$, and $2^i > \sum_{j<i} 2^j$. ☐

Lemma. *If* $M \in \underline{\Lambda}$ *is is weighted with a decreasing weight, and* $M \twoheadrightarrow_{\underline{\beta}} N$ *then* $\mathrm{wt}(N) < \mathrm{wt}(M)$

Proof. Suffices to show it for one step changes. Take Δ the redex in M contracted to get N. Then $\Delta = \underline{(\lambda x.P)Q}$, and as M is decreasing we have for every instance of x in P that $\mathrm{wt}(x) < \mathrm{wt}(Q)$. Thus after the contraction, the weight clearly lowers. □

Lemma. *If $M \in \underline{\Lambda}$ and $M \twoheadrightarrow_{\underline{\beta}} N$ then N is decreasing.*

Proof. Suffices to show for one step changes. Take $\Delta = (\lambda y.A)B$ the redex in M contracted to get N, and take $\Delta_1' = (\lambda x.P')Q'$ in N. Then this comes form a redex $\Delta_1 = (\lambda x.P)Q$ in M. We do cases with how Δ and Δ_1 sit with respect to each other in M. There are five cases: disjoint, Δ_1 in A, Δ_1 in B, Δ in P, Δ in Q.

Case 1: Δ and Δ_1 are disjoint.

Then Δ_1 and Δ_1' are identical, and as Δ_1 was decreasing so is Δ_1'

Case 2: Δ_1 is in A.

We get $\Delta_1' = (\lambda x.P[y := B])(Q[y := B])$. We know that x does not appear in B via alpha conversion. Thus all the x's in P' were in P. And we know that $\mathrm{wt}(x) > \mathrm{wt}(Q)$ for every x in P, and we know that for every y in Q that $\mathrm{wt}(y) > \mathrm{wt}(B)$, and so $\mathrm{wt}(Q[y := B]) < \mathrm{wt}(Q)$, thus $\mathrm{wt}(x) > \mathrm{wt}(Q[y := B])$

Case 3: Δ_1 is in B.

Then Δ_1 only gets copied, so Δ_1' and Δ_1 are identical.

Case 4 Δ is in P.

Then $\Delta_1' = (\lambda x.(\overbrace{\cdots A[y := B] \cdots}^{P'}))Q$. Then any instance of x in P' was in P, and so is weighted higher than Q.

Case 5 Δ is in Q.

Then $\Delta_1' = (\lambda x.P)(\overbrace{\cdots A[y := B] \cdots}^{Q'})$. By the previous lemma, the weight of Q' is smaller than the weight of Q, so we are done. □

Theorem. *$\underline{\beta}$ is strongly normalizing.*

Proof. Take $M \in \underline{\Lambda}$. Weight it with a decreasing ranking. Every step reduces the rank, thus there must be only finitely steps. □

Corollary (Finiteness of Developments)**.** *If $M \in \Lambda$ every development is finite (in fact, every development has length $< 2^{||M||}$ where $||M||$ is the number of variable instances in M)*

Proof. Take $M \in \Lambda$. Consider $\mathscr{F}$ the set of all redexes in M. Lift M to the underlined system, $M^{\mathscr{F}}$. Add a decreasing ranking as in the above proposition where terms are ranked using powers of 2.

Every one step change lowers the rank. Thus the most number of one step reductions is the weight of M itself, which is $\sum_{i=0}^{n-1} 2^i$ where n the number of variable instances in M. This sum is $2^n - 2$. □

1.2 Disjointness

This proof, as far as I know, is mostly due to Micali, Klop, Hyland, and Wadsworth. The ideas were taken from Klop's 1980 PhD thesis.

Definition 3. Fix a term M. Define a relation on subterms $P < Q$ by

- If $P \subseteq Q$ then $P < Q$
- If $\underline{(\lambda x.\cdots P \cdots)(\cdots Q \cdots)}$ then $P < Q$

Call $<^*$ the transitive closure.

Definition 4. If P is a subterm of M and $M \to N$ then we say P' in N is a descendent of P with respect to this reduction if when one labels P in M and does the reduction carrying around the label and resulting in P' being labeled. Here subsitution over a label is defined pretty much how one would expect, except that it is destroyed for a single variable subsitution. Formally we have

$$
\begin{aligned}
x^\alpha[x := N] &= N \\
y^\alpha[x := N] &= y^\alpha \\
(PQ)^\alpha[x := N] &= (P[x := N]Q[x := N])^\alpha \\
(\lambda y.P)^\alpha[x := N] &= (\lambda y.P[x := N])^\alpha \qquad \textit{pick alpha representive } y \neq x \textit{ and avoid capture}
\end{aligned}
$$

The important remark: a residual is a descendent of a redex in M (although not all descendents of a redex are residuals)

Lemma 1. *Suppose $M \to_{\underline{\beta}} N$ and P, Q are subterms of M. If P' and Q' are descendents of P and Q respectively with $P' <^* Q'$ then $P <^* Q$.*

Proof. We do this by induction on the length of $P' <^* Q'$.

First we do the cases where $P' < Q'$.

If it is because $P' \subseteq Q'$ then this is easy as either $P \subseteq Q$ or $\underline{(\lambda x.\cdots P\cdots)(\cdots Q\cdots)}$ where x is in Q. Regardless, $P < Q$.

If it is because $\underline{(\lambda x.\cdots P'\cdots)(\cdots Q'\cdots)}$ then essentially we only have the cases where

$$\underline{(\lambda x.\cdots \underline{(\lambda z.\cdots P\cdots)L}\cdots)(\cdots Q\cdots)}$$

which means $Q < P$ directly if x is still in P, and if x is in L then z is in P and we have $Q < L < P$. The only other case is where

$$\underline{(\lambda z.\cdots \underline{(\lambda x.\cdots P\cdots)(\cdots z\cdots)}\cdots)(\cdots Q\cdots)}$$

But then $Q < \cdots z \cdots < P$. Note, these are essentially the same case, but conceptially one might consider both could happen. Also, the reduction could be disjoint, or just change one of P and Q internally, but those are obvious to handle.

For the transitive case, suppose $P' <^* E < Q'$ then by induction hypothesis $P < E$ and so $P < Q$ ☐

Remark 1. Note that this lemma will be the only place in the proof that we will use that this is a development. If we did not have a development, we would not have been able to say, for instance, that $Q < L < P$ above because the $L < P$ assumes that the redex is marked.

One might argue you could change the definition of $<^*$ to work with a β redex instead of a $\underline{\beta}$ one, but then one would have to do the last lemma with the case where the redex $(\lambda x.\cdots P'\cdots)(\cdots Q'\cdots)$ was created in the one step, and that would lead to problems.

Lemma 2. *If Δ_1 and Δ_2 are redexes and $M \twoheadrightarrow_{\underline{\beta}} N$ and $\Delta_1' \subseteq \Delta_2'$ in N (where Δ_1' and Δ_2' are residuals of Δ_1 and Δ_2 respectively) then $\Delta_1 <^* \Delta_2$.*

Proof. Do induction on the length of the reduction.

If $M = N$ then this is obvious and the reduction is 0 steps, this is obvious.

Otherwise, $M \to M_1 \twoheadrightarrow N$. Δ_1' and Δ_2' are residuals of two redexes, Δ_1'' and Δ_2'' in M_1, which are themselves residuals of Δ_1 and Δ_2. By induction hypothesis, $\Delta_1'' <^* \Delta_2''$. By the last lemma, $\Delta_1 <^* \Delta_2$. ☐

Lemma 3. *If M has a decreasing weight, and $\Delta_1 <^* \Delta_2$ then* $\mathrm{wt}(\Delta_1) < \mathrm{wt}(\Delta_2)$.

Proof. If $\Delta_1 <^* \Delta_2$ because $\Delta_1 \subseteq \Delta_2$ then this is trivial.

If $\Delta_1 <^* \Delta_2$ because $(\lambda x.\cdots\Delta_1\cdots)(\cdots\Delta_2\cdots)$ and x is in Δ_1 this is also easy as the weight of the variable x in Δ_1 has to beat Δ_2.

The transitive case is easy as the weights are natural numbers and $<$ is an ordering on the natural numbers. ☐

Lemma 4. *Suppose M has a decreasing weight and $M \twoheadrightarrow N$ is a development and Δ_1 and Δ_2 are redexes in M with residuals Δ_1' and Δ_2' in N. If $\Delta_1' \subseteq \Delta_2'$ then* $\mathrm{wt}(\Delta_1) < \mathrm{wt}(\Delta_2)$

Proof. Last two lemmas ☐

Corollary 1. *If Δ is a redex in M and $M \twoheadrightarrow_{\underline{\beta}} N$ then all residuals of Δ in N are disjoint.*

Remark 2. One doesn't truely need to use weights here. I just did because the infrastructure was already built, and it gives insight on the connection between these two. One really only needs to show that $<^*$ is a strict order.

A convincing arguement that the ordering is strict without weights is that if $A < B$ then A is either a subterm of B or A is to the left of B. Each of these is strict, and each continues to be strict under it's transitive closure (eg. if $A < B < C < D$ then maybe A is a subterm of B which is to the left of C which is to the left of D, but then A is surely not equal to D since it is to the left of D)

Corollary 2. *In a development, if one sets up all bound variables to be distinct from each other and free variables, one never needs to do an alpha conversion to do a reduction*

Proof. The only reason one would have to do an alpha conversion is to avoid conflicts. A conflict looks like:

$$\cdots(\lambda z.\cdots(\lambda z.\cdots)P\cdots)Q\cdots$$

But here one has residuals of the redex with z as a bound variable is not disjoint from itself. ☐

Remark 3. The last statement is not needed for the proof, it's more of the "bonus" information achieved from doing the proof in this way (like the bound the the number of reductions was a bonus the last way). In fact, one can strengthen it a bit (for free) and omit the "do to a reduction" bit. That is to say, in a development, one will never have any lambda term inside of another lambda term with the same bound variable, not just residuals of redexes. Since I had only been talking about residuals, I thought it was more natural as stated above.

Theorem 2. *$\underline{\beta}$ is strongly normalizing.*

Proof. Label each redex in M with a distinct natural number. For each redex Δ (and eventually each residual) call its color degree the number redexes there are in Δ with different labels. Consider the multiset of color degres.

If $M \twoheadrightarrow N$ then a redex Δ is contracted. If $\Delta' \subseteq \Delta$ then any residual Δ'' of Δ' will have color degree strictly less than that of Δ as Δ'' can not have a internal redex labeled that same as its own label by the corollary; as Δ did have a residual with such a label (namely Δ') in the worst cast Δ'' gets all the other labels that were present in Δ as internal redexes, but this is still less than those in Δ.

Thus the multiset decreases with respect to the usual function between multisets and countable ordinals, so it can only do so finitely many times. ☐

Corollary 3 (Finiteness of Developments)**.** *If $M \in \Lambda$ then every development is finite.*

1.3 Labels

This system is attributed to Hyland and Wadsworth. The proof presented here on its strong normalization is in essense an analog to a proof of strong normalization of the simply typed lambda calculus due to Tait and is attributed by Barendregt to van Daalen.

Definition 5. We define the set of labeled lambda terms Λ^L as follows:

- x a variable, $x \in \Lambda^L$
- $M \in \Lambda^L$ and x a variable then $(\lambda x.M) \in \Lambda^L$.
- $M, N \in \Lambda^L$ then $MN \in \Lambda^L$
- $M \in \Lambda^L$ and $n \in \mathbb{N}$ then $M^n \in \Lambda^L$

Denote $\Lambda^\perp := \Lambda^L \cup \{\perp\}$.

Definition 6. We define a few rewrite rules to replace regular boring β.

- lab: $(M^n)^m \to_l M^{\min\{n,m\}}$

- β_L: $(\lambda x \centerdot M)^{n+1} N \to_{\beta_L} (M[x := N^n])^n$
- $\beta_\bot$: $(\lambda x \centerdot M)^0 N \to_{\beta_\bot} (M[x := \bot])^0$
- $\bot$: $\bot M \to_\bot \bot$ and $\lambda x \centerdot \bot \to_\bot \bot$ and $\bot^n \to_\bot \bot$

Write $M \to_+ N$ to denote the one step reduction under the union of all of those.

Lemma 5. *If M, N are lambda terms and M has no infinite reduction paths and $M[x := N] \twoheadrightarrow \lambda y \centerdot P$ then either $M \twoheadrightarrow x N_1 \cdots N_n$ this $N N_1[x := N] \cdots N_n[x := N] \twoheadrightarrow \lambda y \centerdot P$, or $M \twoheadrightarrow \lambda y \centerdot P'$ and $P'[x := N] \twoheadrightarrow P$.*

Proof. Do induction on $(d(M), ||M||)$ where $d(M)$ is the largest reduction path of M and $||M||$ is the size of M.

Do cases based on the form of M.

If M is a variable and $M = x$ then we're done. Otherwise, if $M \neq x$, $M = z$ but then $M[x := N] = z \neq \lambda y \centerdot P$.

If M is $\lambda y \centerdot N$ then we're done.

If $M = M_1 M_2$ then we have $M_1[x := N] M_2[x := N] \twoheadrightarrow \lambda y \centerdot P$. The head symbol is a lambda, so we must have $M_1[x := N] \twoheadrightarrow \lambda z \centerdot Q$ and

$$(\lambda z \centerdot Q)(M_2[x := N]) \to Q[z := M_2[x := N]] \twoheadrightarrow \lambda y \centerdot P$$

. As $||M_1|| < ||M||$ we can use the induction hypothesis and get $M_1 \twoheadrightarrow \lambda z \centerdot Q'$ and $Q'[x := N] \twoheadrightarrow Q$ or $M_1 \twoheadrightarrow x U_1 \cdots U_n$ and $N U_1[x := N] \cdots U_n[x := N] \twoheadrightarrow \lambda z \centerdot Q$.

In the second case, we are immediately done as $M = M_1 M_2 \twoheadrightarrow x U_1 \cdots U_n M_2$ and

$$N U_1[x := N] \cdots U_n[x := N] M_2[x := N] \twoheadrightarrow (\lambda z \centerdot Q) M_2[x := N] \twoheadrightarrow \lambda y \centerdot P$$

If we have $M_1 = \lambda z \centerdot Q'$ then we have $M \twoheadrightarrow (\lambda z \centerdot Q') M_2$. Then $M \twoheadrightarrow Q'[z := M_2]$ and so

$$M[x := N] \to Q'[z := M_2][x := N] = Q[z := M_2[x := N]] \twoheadrightarrow \lambda y \centerdot P$$

In particular, we have $Q'[z := M_2][x := N] \twoheadrightarrow \lambda y \centerdot P$ and it has a strictly shorter reduction path than M (as a redex was contracted). Applying the induction hypothesis, we get either $Q'[z := M_2]$ reduces to $\lambda y \centerdot P'$ and $P'[x := N] \twoheadrightarrow P$ or it reduces to $x U_1 \cdots U_n$. Regardless, we are done as M would also reduce to these things. □

Lemma 6. *If $(\cdots(M_1^{p_1} M_2)^{p_2} \cdots) M_n)^{p_n} \twoheadrightarrow (\lambda y \centerdot P)^q$ then $q \leq p_i$.*

Proof. First observe that if $Q_1^a \twoheadrightarrow Q_2^b$ then we must have $a \geq b$ as the outer most label can only get smaller by the label contraction rule.

By induction on n. If $n = 0$ then we have $M_1^{p_1} \twoheadrightarrow (\lambda y \centerdot P)^q$, then by the above $q \leq p_1$.

If $m > 0$ then we have, as the entire term reduces to a λ that the first $n-1$ terms reduce to a λ, ie.

$$(\cdots(M_1^{p_1} M_2)^{p_2} \cdots) M_{n-1})^{p_{n-1}} \twoheadrightarrow (\lambda z \centerdot P')^{q'}$$

By induction hypothesis, $q' \leq p_i$ for all i. Then

$$((\lambda z \centerdot P')^{q'} M_n)^{p_n} \twoheadrightarrow ((P'[z := M_n^{q'-1}])^{q'-1})^{p_n} \twoheadrightarrow (\lambda y \centerdot P)^q$$

. Thus the observation and transitivity, $q \leq p_i$ for all i.

□

Lemma 7. *If M is strongly normalizing (with respect to the labeled reductions) then $M[x := \bot]$ is strongly normalizing*

Proof. By induction on $(d(M), ||M||)$. If it's a variable then this is easy. If $M = \lambda y \centerdot P$ then $M[x := \bot] = \lambda y \centerdot P[x := \bot]$ and by induction $P[x := \bot]$ is strongly normalzing, so $\lambda y \centerdot P[x := \bot]$ is as well. If $M = (N)^n$ we can just pass the induction through the label.

If $M = M_1 M_2$ then examine $M_1[x := \bot] M_2[x := \bot]$. Suppose that we had an infinite reduction path. By induction hypothesis, $M_1[x := \bot]$ and $M_2[x := \bot]$ are both strongly normalizing, and so we must have that $M_1[x := \bot]$ reduces to a lambda term. That is, $M_1[x := \bot] \twoheadrightarrow (\lambda y\boldsymbol{.}P)^n$, and $M_2[x := \bot] \twoheadrightarrow Q$ and $(\lambda y\boldsymbol{.}P)^n Q$ has an infinite reduction path by contracting the redex. By the last lemma there are two cases:

Case 1:

$M_1 \twoheadrightarrow (\lambda y\boldsymbol{.}P')^n$ and $P'[x := \bot] \twoheadrightarrow P$.

In the case where the n above is 0, note that $M[x := \bot] \twoheadrightarrow (\lambda y\boldsymbol{.}P)^0 Q \to P[y := \bot]$. $||P|| = ||P'|| < ||M||$ so we can apply the induction hypothesis to P.

Otherwise, $n > 0$. By induction hypothesis, $P'[x := \bot]$ is strongly normalzing. Moreover, M is strongly normalizing so $(P'[x := M_2^{n-1}])^{n-1}$ is as well and has longest reduction path less than M. Thus by induction hypothesis $(P'[y := M_2^{n-1}])^{n-1}[x := \bot]$ strongly normalizing. This reduces to $(P[y := Q^{n-1}])^{n-1}$, which is then a contradiction.

Case 2: $M_1 \twoheadrightarrow xN_1 \cdots N_n$ and $\bot N_1[x := \bot] \cdots N_n[x := \bot] \twoheadrightarrow (\lambda z\boldsymbol{.}P)^n$. But the leftmost symbol will always be $\bot$. This is a contradiction, and completes the proof of the claim. □

Lemma 8. *If M and N are strongly normalizing (with respect to the labeled reductions) then $M[x := N]$ is strongly normalizing.*

Proof. We do this by induction on $(d(M), ||M||, l(N))$ where $l(N)$ is the outer label of N.

If M is a variable, then this is straightforward.

If $M = (P)^n$ this is equally straightforward (just pass past the label).

If $M = \lambda y\boldsymbol{.}P$ then $M[x := N] = \lambda y\boldsymbol{.}P[x := N]$. As $||P|| < ||M||$ we are done by induction hypothesis.

The only difficult case is when $M = M_1 M_2$. As M is strongly normalizing, so much M_1 and M_2 be, and thus by induction hypothesis, $M_1[x := N]$ and $M_2[x := N]$ are also strongly normalizing. Assume for sake of contradiction that $M_1[x := N]M_2[x := N]$ is not strongly normalizing. As each component is, the only way this could happen is if $M_1[x := N]$ reduces to a term which is not neutral. Then $M_1[x := N] \twoheadrightarrow (\lambda y\boldsymbol{.}P)^n$ and $M_2[x := N] \twoheadrightarrow Q$ and $(\lambda y\boldsymbol{.}P)^n Q$ must have an infinite reduction path which contracts that redex.

By the lemma above, we distinguish cases:

Case 1: $M_1 \twoheadrightarrow (\lambda y\boldsymbol{.}P')^n$ and $P'[x := N] \twoheadrightarrow P$

In this case if $n = 0$ then $M_1 M_2 \twoheadrightarrow (\lambda y\boldsymbol{.}P')^0 M_2$. $P = P'[x := N]$, and by the above lemma $(P'[x := N][y := \bot])^0$ is strongly normalizing.

Otherwise, we have $M_1 M_2 \twoheadrightarrow (\lambda y\boldsymbol{.}P')^{l+1} M_2 \to (P'[y := M_2^l])^l$ Here, we can use the induction hypothesis as there is a shorter longest reduction path. Thus $(P'[y := M_2^l])^l[x := N]$ is strongly normalizing. But

$$(P'[y := M_2^l])^l[x := N] = P[y := M_2^l[x := N]]^l \twoheadrightarrow P[y := Q^l]^l$$

which is a contradiction as this was assumed to not be strongly normalizing.

Case 2: $M_1 \twoheadrightarrow (x^{l_1} N_1)^{l_2} \cdots N_k)^{l_k}$ and $(x^{l_1} N_1)^{l_2} \cdots N_k)^{l_k}[x := N] \twoheadrightarrow (\lambda y\boldsymbol{.}P)^n$

Then n must be smaller than the each of the labels on this term; in particular, n is smaller than the label on N. Thus, as P is strongly normalizing, I can substitute any label $< n$ into it. Thus $(P[x := Q^{n-1}])^{n-1}$ is strongly normaizing, which is a contradiction. □

Theorem 3. *Labelled reductions are strongly normalizing.*

Proof. By induction on term. If M is a variable, this is trivial (variables don't reduce so much).

If $M = \lambda x\boldsymbol{.}P$ then we just pass thru the induction hypothesis to P. Similar if M is a term with an outer label.

If $M = M_1 M_2$ and M did have an infinite reduction path, then $M_1 \twoheadrightarrow \lambda y\boldsymbol{.}P$ and $M_2 \twoheadrightarrow Q$ and $(\lambda y\boldsymbol{.}P)^n Q$ will have an infinite reduction path through the contraction of that redex. If $n = 0$ then $(P[y := \bot])^0$ is strongly normalizing by the lemma. Otherwise, $n > 0$, and we have $P[y := Q^{n-1}]^{n-1}$ is not strongly normalizing, but as P and Q both are, by the last lemma, it is, which is a contradiction. □

Theorem 4 (Finiteness of Developments)**.** *All developments are finite*

Proof. Take M to be a term and $\mathscr{F}$ a set of redexes in M. Consider just the notion of reduction l. Label all the abstraction terms in each of the redexes in $\mathscr{F}$ with label 1. Then any l reduction will only contract those redexes and residuals. By last theorem, such a reduction is finite. □

2 Church Rosser

Knowing that developments are finite gives us a proof of the Church-Rosser theorem (also known as confluence) very quickly.

Definition 7. For a rewrite system with reduction rule $\to$ we say it has the diamond property if for every M, M_1, M_2 such that $M \to M_1$ and $M \to M_2$ there exists a M_3 such that $M_1 \to M_3$ and $M_2 \to M_3$. (I read this as "we can fix small mistakes quickly")

We say it is Church-Rosser, or confluent, if the transitive closure of $\to$, denoted by $\twoheadrightarrow$, has the diamond property. (I read this "we can fix mistakes")

We say it is Weak Church-Rosser, or weakly confluent, if for every M, M_1, M_2 such that $M \to M_1$ and $M \to M_2$ there exists a M_3 such that $M_1 \twoheadrightarrow M_3$ and $M_2 \twoheadrightarrow M_3$. (I read this as "we can fix small mistakes")

Definition 8. Call a development complete if it contracts all of the underlined redexes.

Lemma 9. *$\underline{\beta}$ is weak church rosser, that is to say if $M \to_{\underline{\beta}} N_1$ and $M \to_{\underline{\beta}} N_2$ then there is a N such that $N_1 \twoheadrightarrow_{\underline{\beta}} N$ and $N_2 \twoheadrightarrow_{\underline{\beta}} N$.*

Proof. Let $\Delta_1 = (\lambda x\,.\,P_1)Q_1$ be the redex contracted in M to get N_1 and $\Delta_2 = (\lambda y\,.\,P_2)Q_2$ the redex contracted in M to get N_2. Just do cases on how these two redexes sit with respect to each other.

If they are disjoint then

$$N_1 = \cdots P_1[x := Q_1] \cdots \underline{(\lambda y\,.\,P_2)Q_2} \cdots$$
$$N_2 = \cdots \underline{(\lambda y\,.\,P_1)Q_1} \cdots P_2[y := Q_2] \cdots$$

then join them by $N = P_1[x := Q_1] \cdots P_2[y := Q_2] \cdots$

If Δ_2 sits in P_1 then

$$N_1 = \cdots \underline{(\lambda x\,.\,\cdots P_2[y := Q_2] \cdots)Q_1} \cdots$$
$$N_2 = \cdots (\cdots \underline{\lambda y\,.\,(P_2[x := Q_1])(Q_2[x := Q_1])} \cdots) \cdots$$

Then join them by $N = \cdots(\cdots P_2[x := Q_1][y := Q_2[x := Q_1]] \cdots) \cdots$, which is the same as $N = \cdots(\cdots P_2[y := Q_2][x := Q_1] \cdots) \cdots$.

If Δ_2 sits in Q_1 then

$$N_1 = \cdots \underline{(\lambda x\,.\,P_1)(\cdots P_2[y := Q_2] \cdots)} \cdots$$
$$N_2 = \cdots P_1[x := \cdots \underline{(\lambda y\,.\,P_2)Q_2} \cdots] \cdots$$

Then join them with $N = \cdots P_1[x := \cdots P_2[y := Q_2] \cdots]$. Note: unlike in the other cases, here may need to do more than one reduction in N_2 since that Δ_2 redex may have been copied many times. ☐

Remark 4. We used nothing about underlining above. That is, in fact, a proof that β itself is weakly Church-Rosser if one erases all the underlines.

Lemma 10 (Newman's Lemma)**.** *If a system of reduction is strongly normalizing and weakly Church-Roser then it is Church Rosser.*

Proof. Do it by induction of the longest reduction path. If it is 0, then there is not much to check.

Take a term M and suppose $M \twoheadrightarrow N_1$ and $M \twoheadrightarrow N_2$. Then $M \to M_1 \twoheadrightarrow N_1$ and $M \to M_2 \twoheadrightarrow N_2$. Then by weak Church rosser $M_1 \twoheadrightarrow P$ and $M_2 \twoheadrightarrow P$. But the longest reduction path in M_1 is shorter than that of M, and similarly for M_2. Thus there is a Q_1 such that $N_1 \twoheadrightarrow Q_1$ and $P \twoheadrightarrow Q_1$ and a Q_2 such that $N_2 \twoheadrightarrow Q_2$ and $P \twoheadrightarrow Q_2$.

Then $P \twoheadrightarrow Q_1$ and $P \twoheadrightarrow Q_2$, so there is a N such that $Q_1 \twoheadrightarrow N$ and $Q_2 \twoheadrightarrow N$. By then $N_1 \twoheadrightarrow Q_1 \twoheadrightarrow N$ and $N_2 \twoheadrightarrow Q_2 \twoheadrightarrow N$. (drawing a picture helps!) ☐

Lemma 11. *$\underline{\beta}$ is Church Rosser.*

Proof. It is strongly normalizing by the last section. The last two results complete the proof. ☐

Theorem 5 (The Uniqueness of Complete Developments)**.** *If $\mathscr{F}$ is a set of redexes in M then there exists exactly one N such that $M^{\mathscr{F}} \twoheadrightarrow_{\underline{\beta}} N$ and N has no underlined terms.*

Proof. The existence is from developments being finite, so one can be produced by an reduction approach you like.

The uniqueness comes from Church-Rosser; if one could reduce to two such terms N_1 and N_2 then there would have to exist a N that each reduce to. But as they have no underlined terms, they can not $\underline{\beta}$ reduce. Thus they must be equal. □

Definition 9. We define a new form of reduction $\rightarrow^*$ on Λ. We say $M \rightarrow^* N$ if there is some set of redexes in M, call it $\mathscr{F}$, such that N is the term obtained by doing a complete development on $M^{\mathscr{F}}$

Lemma 12. *$\rightarrow^*$ has the diamond property. This is to say, if $M \rightarrow^* N_1$ and $M \rightarrow^* N_2$ then there is a N such that $N_1 \rightarrow^* N$ and $N_2 \rightarrow^* N$*

Proof. Take $\mathscr{F}_1$ and $\mathscr{F}_2$ witnesses to $M \rightarrow^* N_1$ and $M \rightarrow^* N_2$ respectively. Let $\mathscr{F} = \mathscr{F}_1 \cup \mathscr{F}_2$. Let N be the complete development of $M^{\mathscr{F}}$.

It is obvious that N is attainable from N_1 and N_2. To see this more clearly, do a development of $M^{\mathscr{F}}$ that contracts redexes from $\mathscr{F}_1$ first. You will eventually arrive at N_1 with some underlined redexes. These are exactly the redexes one must underline to get to N from N_1. Thus $N_1 \rightarrow^* N$. By symmetry, the same holds for N_2. □

Theorem 6 (Church Rosser for β)**.** *Λ with β reduction is Church-Rosser*

Proof. View a one step reduction of a redex Δ as a complete development of $\{\Delta\}$. Then if $M \rightarrow P_1 \rightarrow \cdots \rightarrow P_k \rightarrow N_1$ and and $M \rightarrow Q_1 \rightarrow \cdots \rightarrow Q_l \rightarrow N_2$, one can convert this to $M \rightarrow^* P_1 \rightarrow^* \cdots \rightarrow^* P_k \rightarrow^* N_1$ and and $M \rightarrow^* Q_1 \rightarrow^* \cdots \rightarrow^* Q_l \rightarrow^* N_2$. Using the last lemma, we can diagram chase to an N that N_1 and N_2 both reduce to. Then it's just a matter of reading off a β reduction. □