Some Basic Operations on Languages

(1) Union : $\boldsymbol{L = L_1 \cup L_2}$

(2) Concatenation : $\boldsymbol{L = L_1 \, . \, L_2}$

(3) Closure (star or Kleene closure) $\boldsymbol{L^* = \cup_{k=0,\infty} L^k}$

Definition of a set RE of regular expressions

(over a finite set $\Sigma := \{\sigma_1, \sigma_2, \ldots \sigma_K\}$ *)*

Recursive Formal Definition

(A) *(Basis)* ***e*** *,* $\varnothing$ *and* $\sigma_1, \sigma_2, \ldots, \sigma_K$ *are all elements of* ***RE***

(B) *(Recursion)*

(1) If ***E*** *and* ***F*** *are in* ***RE*** *then so is* ***E+F***

(2) If ***E*** *and* ***F*** *are in* ***RE*** *then so is* ***E.F***

(3) If ***E*** *is in* ***RE*** *then so is* ***E****

(4) If ***E*** *is in* ***RE*** *then so is* ***(E)***

We call each element of the set ***RE*** *a* ***regular expression*** *!*

Language interpretation is a mapping $L : RE \rightarrow 2^{\Sigma^*}$ given by :

$L(e) := \{e\}$ *where* e : = *empty string*

$L(\varnothing) := \varnothing$ *where* $\varnothing$: = *null language (with no strings)*

$L(\alpha) := \{\alpha\}$, $L(\beta) := \{\beta\}$, ... *etc.*

$L(E + F) := L(E) \cup L(F)$

$L(E.F) := L(E).\ L(F)$

$L(E^*) := L(E)^*$

$L((E)) := (L(E))$

Definition *: A language L is called a* ***regular language*** *if it is the language interpretation of a* ***regular expression***

Main Theorem

A language is regular if and and only if it is accepted by some finite state automaton

Proof of the Main Theorem

(if) **Idea :**

(1) Given a DFA $\mathbf{D = (Q, \Sigma, \delta, 1, F)}$ *with* $\mathbf{Q = \{1,2, \ldots, n\}}$

(2) Let $\mathbf{R_{ij}^{\,k}}$ *denote the language corresponding to all paths of* $\mathbf{D}$ *that start at state* $\mathbf{i}$ *; end at state* $\mathbf{j}$ *; and visit intermediate states with numbers* $\mathbf{p \le k}$

(3) Note that $\mathbf{L(D) = \cup_{(m \in F)} R_{1m}^{\,n}}$ *where* $\mathbf{1}$ *is the initial state*

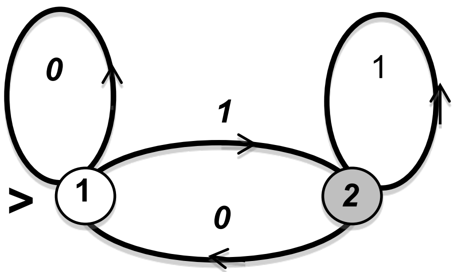
(4) Prove by induction on $\mathbf{k}$ *that* $\mathbf{R_{ij}^{\,k}}$ *is a* **RE** *for all* $\mathbf{i,j = 1,\ldots,n}$ *and* $\mathbf{k = 0,\ldots, n}$. *(see the next slide first formula)*

(5) Conclude that $\mathbf{L(D)}$ *is a* **RE**

The Inductive Formula for DFA → RE

$$R_{ij}^{k} = R_{ij}^{k-1} + R_{ik}^{k-1}.\ (R_{kk}^{k-1})^{*}.\ R_{kj}^{k-1} \quad ;\ i,j=1,\ldots,n\ ;\ k=0,\ldots,n$$

Example



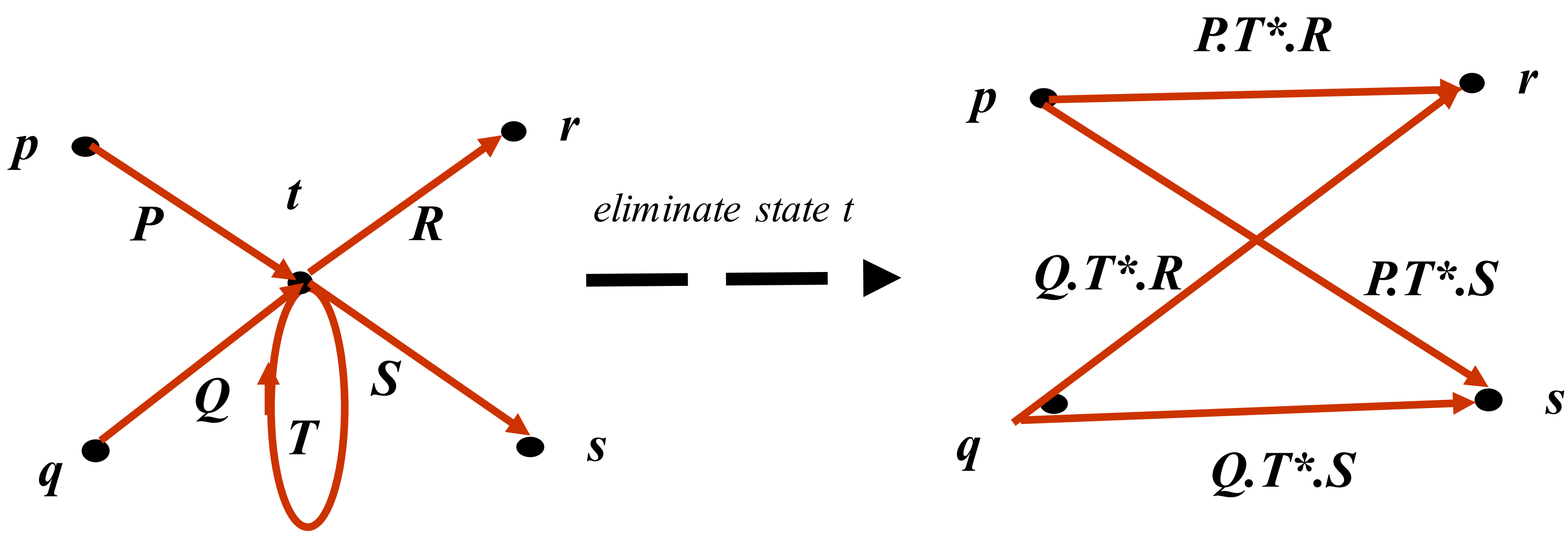
$$R_{11}^{0} = 0+e\ ;\ R_{22}^{0} = 1+e\ ;\ R_{21}^{0} = 0\ ;\ R_{12}^{0} = 1$$

$$R_{11}^{1} = 0^{*}\ ;\ R_{22}^{1} = 1+0.0^{*}.1+e\ ;\ R_{21}^{1} = 0.0^{*}\ ;\ R_{12}^{1} = 0^{*}.1$$

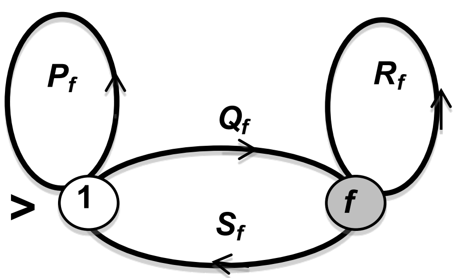
$$R_{11}^{2} = \ldots\ ;\ R_{22}^{2} = \ldots\ ;\ R_{21}^{2} = \ldots\ ;\ R_{12}^{2} = \ldots$$

After Simplification : $L=R_{12}^{2} = (0^{*}.1.1^{*}.0)^{*}.0^{*}.1.1^{*}$

Alternative Proof of the Main Theorem (State Elimination)



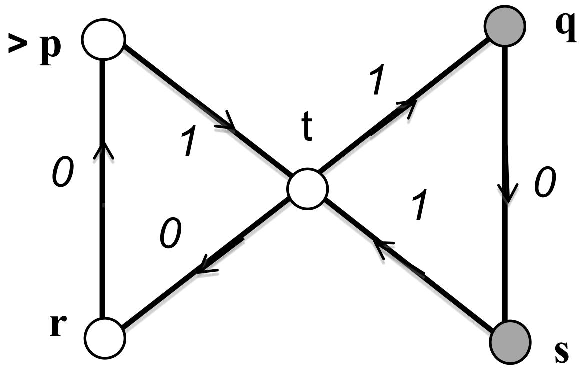
After eliminating all non-initial and non-final states ; start eliminating all final states except one ***f*** *in* ***F*** *and repeat this for each distinct* ***f*** *in* ***F***. *Then the following picture(s) prevail*



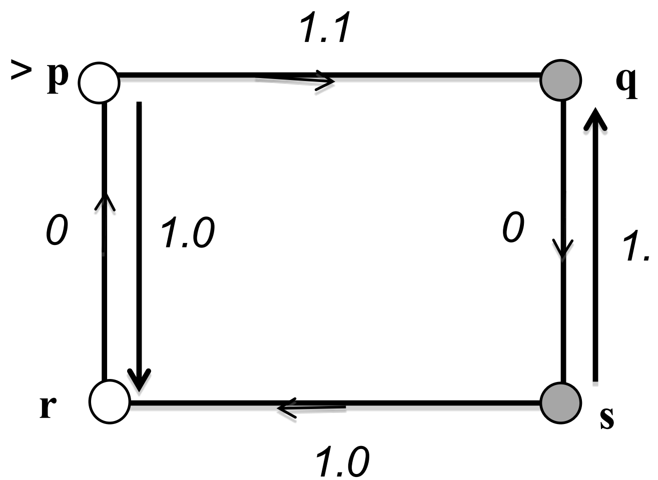
$$L_f = (P_f^*.Q_f.R_f^*.\ S_f)^*.P_f^*.\ Q_f.\ R_f^*$$

$$L = \Sigma_{(f \in F)}\ L_f$$

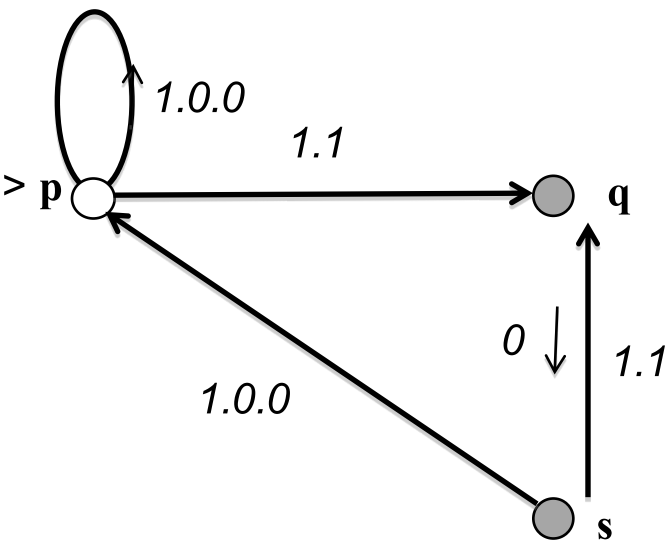
Example



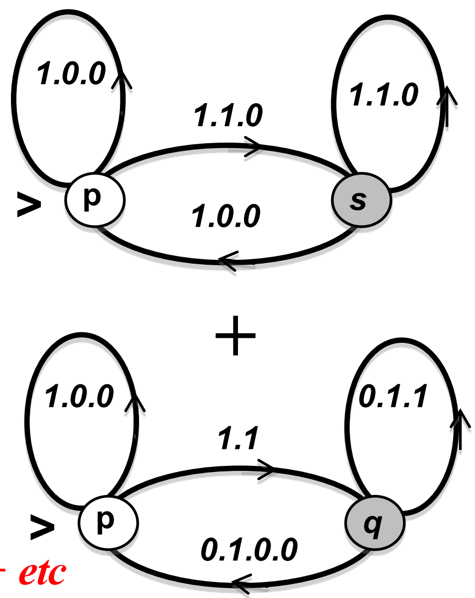
eliminate t



eliminate r



eliminate q and s



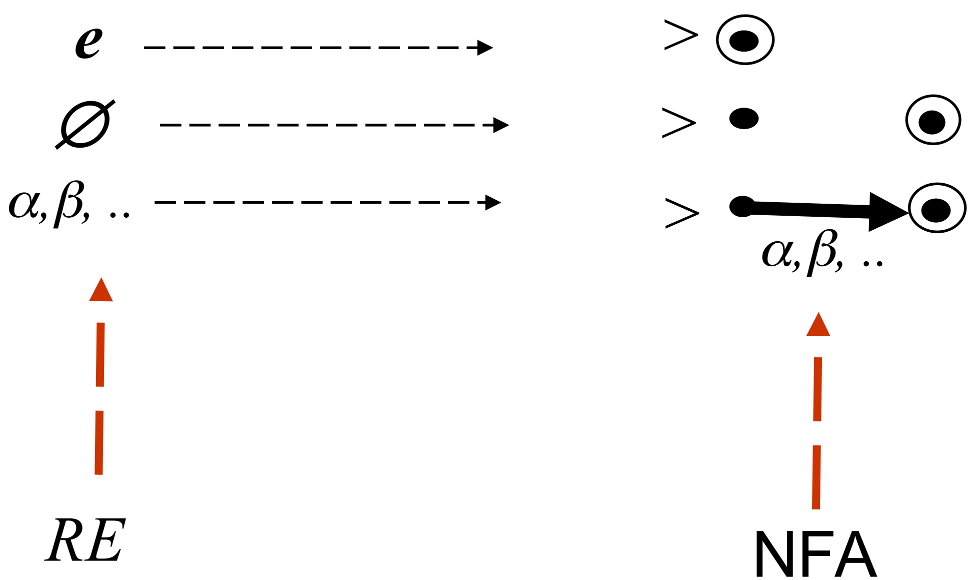
L_f=((1.0.0)*.(1.1.0).(1.1.0)*.(1.0.0))*.(1.0.0)*.(1.1.0).(1.1.0)* + etc

Proof of the Main Theorem

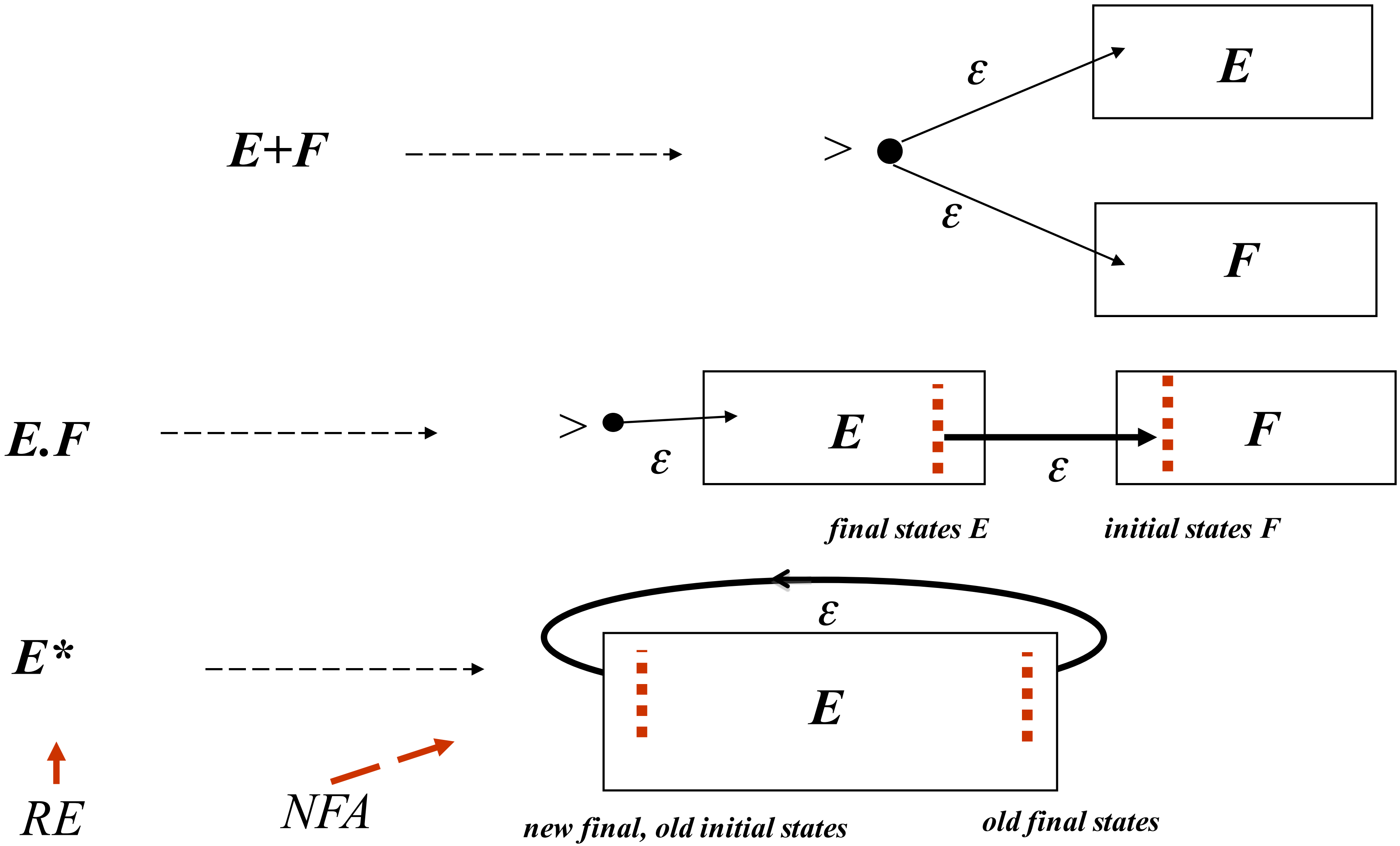
(only if) ***Idea :***

given ***RE****s over the set* $\boldsymbol{\Sigma = (\alpha, \beta, \gamma, \ldots,)}$

Basis

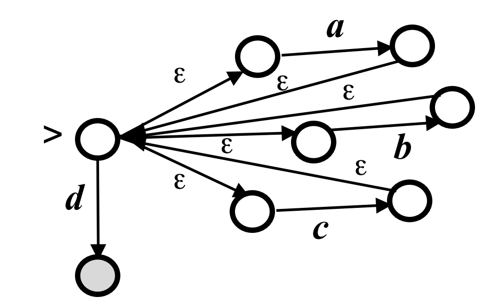


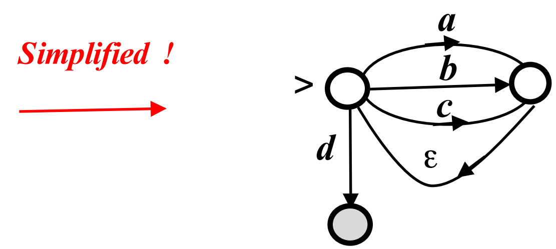
Proof of the **Main Theorem** (continued) (only if)



Some short cuts !

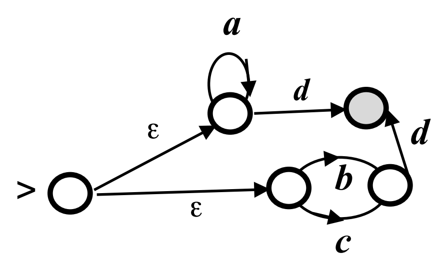
$(a+b+c)^*.d$

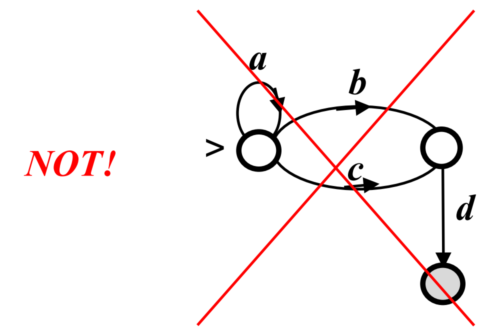




But !

$(a^*+b+c).d$





Algebraic Laws For REs

Trivial Laws

(1) $L+M = M+L$; $(L+M) + N = L + (M+N)$; $(L.M).N = L.(M.N)$

(2) $\phi + L = L$; $e.L = L.e = L$; $\phi.L = \phi$

(3) $L.(M+N) = L.M + L.N$; $(L+M).N = L.N + M.N$; $L+L = L$

Non-trivial Laws

(4) $(L+M)^* = (L^*+M^*)^* = (L^*.M^*)^*$

(5) $(L.M)^* \subseteq (L^*.M^*)^*$ and $(L.M)^* = (L^*.M^*)^*$ iff $e \in L$ **and** $e \in M$

Proof of $(L+M)^* = (L^*.M^*)^*$

Two steps : (1) $(L+M)^* \subseteq (L^*.M^*)^*$ *; (2)* $(L^*.M^*)^* \subseteq (L+M)^*$

(1) Let $u \in (L+M)^*$ *then* $u = u_1 . u_2 . \ldots . u_k$ for some integer $k \geq 0$ where for each j , $u_j \in L+M$;

but $L \subseteq L^* \subseteq L^*. e \subseteq L^* . M^*$ *and* $M \subseteq M^* \subseteq e . M^* \subseteq L^*. M^*$ *;*

hence $u_j \in L^*. M^* + L^*. M^* = L^*. M^*$ and therefore $(L+M)^* \subseteq (L^* . M^*)^*$

(2) Conversely let $u \in (L^*.M^*)^*$ *then by definition* $u = u_1 . u_2 . \ldots . u_k$ *where* $u_j \in L^*.M^*$ *;*

hence $u_j = v_j^1 . v_j^2 \ldots . v_j^{l(j)} . w_j^1 . w_j^2 \ldots . w_j^{p(j)}$ *where* $v_j^m \in L$ and $w_j^m \in M$ *;*

thus $u = z_1 . z_2 . \ldots z_q$ *where* $q = \sum_{j=1,k} l(j)+p(j)$ *and each* $z_i \in L+M$.

this proves that $(L^*.M^*)^* \subseteq (L+M)^*$

Proof of $(L+M)^* = (L^*+M^*)^*$ given (2) $\rightarrow (L+M)^* = (L^*. M^*)^*$

Since $L \subseteq L^*$ and $M \subseteq M^*$ it follows that $(L+M)^* \subseteq (L^*+M^*)^*$

Conversely let $u \in (L^*+M^*)^*$ then $u = (v_1+w_1). \ \ldots \ . (v_k + w_k)$ where for each j $v_j \in L^*$ and $w_j \in M^*$.

We show that $u \in (L^*. M^*)^*$ by using induction on k.

For $k=1$ $\quad v_1 \in L^* \subseteq L^*. e \subseteq L^*.M^* \subseteq (L^*. M^*)^*$

similarly $w_1 \in M^* \subseteq e. M^* \subseteq L^*.M^* \subseteq (L^*. M^*)^*$ hence $v_1+w_1 \subseteq (L^*. M^*)^*$.

Now assume statement holds for $k-1$, hence $z := (v_1+w_1). \ \ldots \ . (v_{k-1} + w_{k-1}) \in (L^*. M^*)^*$

But using the above reasoning for v_1+w_1 it follows that $v_k+w_k \in (L^*. M^*)^*$

and therefore $u = z . (v_k+w_k) \in (L^*. M^*)^*. (L^*. M^*)^* = (L^*. M^*)^*$ using the obvious

identity $K^* . K^* = K^*$ for any language K. This proves that $(L^*+M^*)^* \subseteq (L^*. M^*)^*$

but by (2) $(L+M)^* = (L^*. M^*)^*$ hence $(L^*+M^*)^* \subseteq (L + M)^*$ and result follows