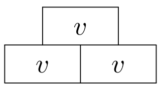
REPRESENTING DOUBLE-INTERPRETATIONS

Markku Laine

Department of Mathematics and Turku Centre for Computer Science,
University of Turku
Turku, 20014, Finland
majlain@utu.fi

In this paper a new notation for representing double-interpretations is introduced. Especially when working with word equations, one often needs to look at two factorizations or interpretations of a word. This usually leads to a complicated case analysis, see for example [3]. To make these kind of considerations easier to the reader, one traditionally provides some kind of tile or line-arc pictures, see [1, 2, 4]. The problem is that these pictures are not exact enough and might not handle all the cases. Hence, additional information has to be added to the base text. The new notation is intended to counteract these difficulties, by providing a formally defined, exact, and also more compact way of representing this information, but still maintaining the illustrative nature of a picture.

As an example the fact that a word v is imprimitive is equivalent to the fact that v is a proper factor of vv, which is represented by the following tile picture.



Same thing is represented by the formula

$$+\frac{][}{[\quad v \quad]} * \frac{[\quad v \quad]}{][} * \frac{][}{[\quad v \quad]} + . \qquad (1)$$

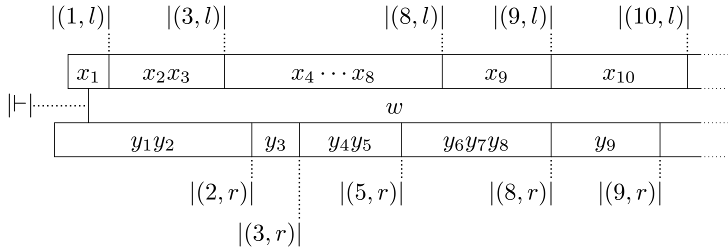
If we would replace the symbols $+$ in the beginning and the end of the formula with symbols $*$, then the cases that v is a prefix and that v is a suffix of vv would also be included. The detailed definition is given next.

Assume that a word w has two interpretations $x_1x_2\cdots x_m$ and $y_1y_2\cdots y_n$, that is

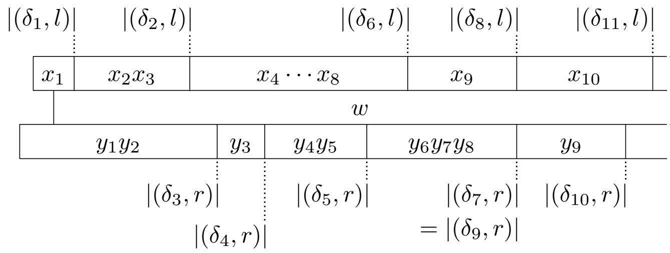
$$p_1wq_1 = x_1x_2\cdots x_m \qquad \text{and} \qquad p_2wq_2 = y_1y_2\cdots y_n, \qquad (2)$$

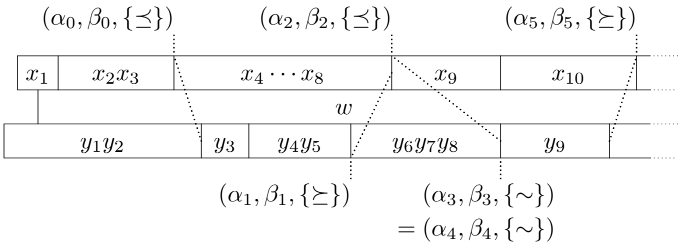
where $|p_1| < |x_1|$, $|q_1| < |x_m|$, $|p_2| < |y_1|$ and $|q_2| < |y_n|$, with $f_i = |x_1\cdots x_i| - |p_1|$, for $i = 0,\ldots,m$; and $g_j = |y_1\cdots y_j| - |p_2|$, for $j = 0,\ldots,n$. A *restriction* is an explicitly given relation $(<, \leq, =, \geq$ or $>)$ between two numbers f_i and g_j, that is, a relation *we are aware of.*

Before the main definition, we need some auxiliary definitions. Example pictures are provided along the way in order to illustrate the concepts in these definitions. We call a pair $(\delta, \epsilon) \in \big(\{0, \dots, m\} \times \{l\}\big) \cup \big(\{0, \dots, n\} \times \{r\}\big)$ a *factor point*, or more precisely a *left factor point*, if $\epsilon = l$, or a *right factor point*, if $\epsilon = r$. Moreover, $\vdash$ is called the *starting factor point* and $\dashv$ the *ending factor point*. The *length* $|(\delta, \epsilon)|$ of a factor point (δ, ϵ) equals f_δ, if $\epsilon = l$, or g_δ, if $\epsilon = r$; and $|\vdash| = 0$, $|\dashv| = |w|$. The length concept allows us to write for any two factor points φ and ψ that $\varphi \prec \psi$, $\varphi \preceq \psi$, $\varphi \sim \psi$, $\varphi \succeq \psi$ or $\varphi \succ \psi$, if we have the restriction $|\varphi| < |\psi|$, $|\varphi| \leq |\psi|$, $|\varphi| = |\psi|$, $|\varphi| \geq |\psi|$ or $|\varphi| > |\psi|$, respectively.



Assume that $P = \big\{(\delta_1, \epsilon_1), \dots, (\delta_\ell, \epsilon_\ell)\big\}$ is a chain of factor points, written in non-decreasing order, that is, for any two factor points (δ_i, ϵ_i) and (δ_j, ϵ_j), with $i < j$, the relation $(\delta_i, \epsilon_i) \preceq (\delta_j, \epsilon_j)$ holds. Here we assume that $\vdash$ and $\dashv$ are not included in P, but allow P to be a multiset. Moreover, we assume that $\vdash$ and $\dashv$ are comparable with all the factor points in P. We call a triple (α, β, Γ) a *constraint of* P, if (α, l) and (β, r) are two consecutive factor points of P, that is, if $\big\{(\alpha, l), (\beta, r)\big\} = \big\{(\delta_i, \epsilon_i), (\delta_{i+1}, \epsilon_{i+1})\big\}$, for some $i \in \{1, \dots, \ell - 1\}$; and $\Gamma = \big\{\gamma \in \{\prec, \preceq, \sim, \succeq, \succ\} \mid (\alpha, l)\ \gamma\ (\beta, r)\big\}$. Moreover, constraint (α, β, Γ) is called *even*, if $\sim \in \Gamma$; *negative*, if it is not even and $\preceq \in \Gamma$; *positive*, if it is not even and $\succeq \in \Gamma$; *inner*, if $\vdash \preceq (\alpha, l) \preceq \dashv$ and $\vdash \preceq (\beta, r) \preceq \dashv$; *outer*, if it is not inner. It follows from the construction that $\max\{f_{\alpha_1}, g_{\beta_1}\} \leq \min\{f_{\alpha_2}, g_{\beta_2}\}$ or $\max\{f_{\alpha_2}, g_{\beta_2}\} \leq \min\{f_{\alpha_1}, g_{\beta_1}\}$ holds, for any two constraints $(\alpha_1, \beta_1, \Gamma_1) \neq (\alpha_2, \beta_2, \Gamma_2)$ of P. Hence, we get an order for constraints.





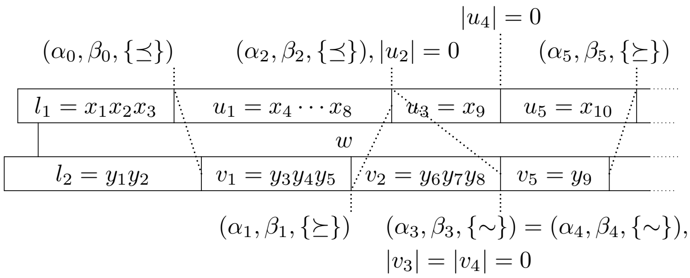
Assume that $C=\{(\alpha_0,\beta_0,\Gamma_0),\ldots,(\alpha_k,\beta_k,\Gamma_k)$ is the set of all constraints of P, written in non-decreasing order. We set $l_1=x_1\cdots x_{\alpha_0}$, $l_2=y_1\cdots y_{\beta_0}$, $u_i=x_{\alpha_{i-1}+1}\cdots x_{\alpha_i}$, $v_i=y_{\beta_{i-1}+1}\cdots y_{\beta_i}$, $r_1=x_{\alpha_k+1}\cdots x_m$, $r_2=y_{\beta_k+1}\cdots y_n$, for $i=1,\ldots,k$; and overlaps

$$s_j=\begin{cases}(p_1^{-1}l_1u_1\cdots u_j)^{-1}(p_2^{-1}l_2v_1\cdots v_j) & \text{if } \vdash\,\preceq(\alpha_j,l)\preceq(\beta_j,r)\preceq\,\dashv,\\ (p_2^{-1}l_2v_1\cdots v_j)^{-1}(p_1^{-1}l_1u_1\cdots u_j) & \text{if } \vdash\,\preceq(\beta_j,r)\preceq(\alpha_j,l)\preceq\,\dashv,\end{cases}$$

for $j=0,\ldots,k$. Now, equalities (2) can be rewritten as

$$p_1wq_1=l_1u_1\cdots u_kr_1 \qquad \text{and} \qquad p_2wq_2=l_2v_1\cdots v_kr_2. \tag{3}$$

It follows from the construction, that if $(\alpha_{i-1},\beta_{i-1},\Gamma_{i-1})$ and $(\alpha_i,\beta_i,\Gamma_i)$ are inner constraints, i. e., the pair (u_i,v_i) is inside w, then u_i and v_i overlap, that is $f_{\alpha_i}\geq g_{\beta_{i-1}}$ and $g_{\beta_i}\geq f_{\alpha_{i-1}}$.



Finally we are ready for the main definition. We depict the constraints in C as a formula

$$\circ_0\;\rho_1\;\circ_1\cdots\circ_{k-1}\;\rho_k\;\circ_k\;, \tag{4}$$

which we call a *representation formula of double-interpretation.* In this formula the symbols $\circ_i$ are operators and ρ_j are terms defined as follows. Operator $\circ_i$ equals one of the following: $+$ or $+s_i$, if s_i is not empty ($\{\prec,\succ\}\cap\Gamma\neq\emptyset$); $*$ or $*s_i$, if s_i might be empty ($\{\prec,\sim,\succ\}\cap\Gamma=\emptyset$); or empty (left out), if s_i is known to be empty ($\sim\in\Gamma$). Moreover, operators $+$ and $*$ are replaced with $\oplus$ and $\circledast$, respectively, if $(\alpha_i,\beta_i,\Gamma_i)$ is an outer constraint. Next, we define the terms ρ_j. If both $(\alpha_{j-1},\beta_{j-1},\Gamma_{j-1})$ and $(\alpha_j,\beta_j,\Gamma_j)$ are negative (resp. positive), then

$$\rho_j=\frac{\uparrow u_j\uparrow}{\downarrow v_j\downarrow}\qquad\left(\text{resp. }\frac{\uparrow u_j\uparrow}{\downarrow v_j\downarrow}\right).$$

If $(\alpha_{j-1}, \beta_{j-1}, \Gamma_{j-1})$ is positive and $(\alpha_j, \beta_j, \Gamma_j)$ is negative (resp. $(\alpha_{j-1}, \beta_{j-1}, \Gamma_{j-1})$ is negative and $(\alpha_j, \beta_j, \Gamma_j)$ is positive), then

$$\rho_j = \frac{\upharpoonright u_j \upharpoonright}{\downharpoonleft \quad v_j \quad \downharpoonleft} \qquad \left(\text{resp. } \frac{\upharpoonright \quad u_j \quad \upharpoonright}{\downharpoonleft v_j \downharpoonleft}\right).$$

If $(\alpha_{j-1}, \beta_{j-1}, \Gamma_{j-1})$ (resp. $(\alpha_j, \beta_j, \Gamma_j)$) is even, then

$$\rho_j = \frac{\upharpoonright u_j \upharpoonright}{\downharpoonleft \quad v_j \quad \downharpoonleft} \quad \text{or} \quad \frac{\upharpoonright \quad u_j \quad \upharpoonright}{\downharpoonleft v_j \downharpoonleft} \quad \left(\text{resp. } \frac{\upharpoonright \quad u_j \quad \upharpoonright}{\downharpoonleft v_j \downharpoonleft} \quad \text{or} \quad \frac{\upharpoonright u_j \upharpoonright}{\downharpoonleft \quad v_j \quad \downharpoonleft}\right)$$

depending on whether $(\alpha_j, \beta_j, \Gamma_j)$ (resp. $(\alpha_{j-1}, \beta_{j-1}, \Gamma_{j-1})$) is negative or positive. To conclude the definition, the end markers $\upharpoonright \ldots \upharpoonright$ of the upper part of the term represent square brackets equal to $[\ldots]$, if u_j is known to be nonempty; $]\ldots[$, if u_j might also be empty. Moreover, u_j is omitted and brackets written near each other ( ][ ) if u_j is known to be empty. The lower part $\downharpoonleft \ldots \downharpoonleft$ is set similarly with respect to v_j.

The operators $+s_i$ or $*s_i$ are used in the case we need to explicitly denote the overlap s_i. If all the constraints are inner, then the formula actually represents a double-factorization. If any of the words l_i or r_i is nonempty, then the representation is only partial, that is, it represents some middle part of the two interpretations or factorizations.

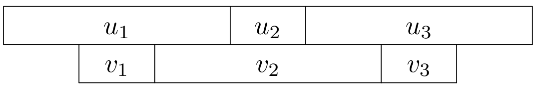
Observe that a pair of interpretations of a word can have several different representation formulas. Also many pairs of interpretations of words can have the same representation formula. Usually we only use the representation formulas without explicitly writing down the two interpretations, the set of factor points or the set of constraints.

To better illustrate the intuition behind the definition, an example of a five step transformation from a traditional tile picture of words to representation formula is given next.

Step 1. Assume that we have an equality

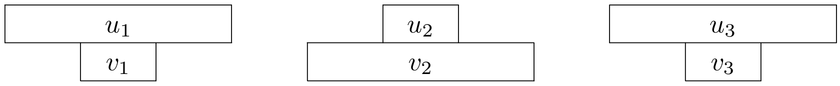
$$l_1 u_1 u_2 u_3 r_1 = l_2 v_1 v_2 v_3 r_2$$

represented as a picture



where we already excluded the words l_i and r_i. This picture tells us, for example, that v_1 is a factor of u_1 and that u_1 overlaps with v_2. It does not tell us, however, whether v_1 is or can be empty or not or whether the overlap word of u_1 and v_2 is empty, i.e., whether u_1 and v_2 are only neighbouring each other.

Step 2. As a second step, we rip mercilessly the previous picture into three separate blocks.



Step 3. Next we transform the blocks into terms of representation formula.

$$\frac{[\quad u_1 \quad]}{[v_1]} \qquad \frac{][}{[\quad v_2 \quad]} \qquad \frac{[\quad u_3 \quad]}{]v_3[}$$

We can now read, that u_2 is empty (it is, in fact, left out), v_3 might be empty and all the rest of the words are nonempty.

Step 4. This time we add the operators before, in between and after the terms.

$$+\frac{[\quad u_1 \quad]}{[v_1]} * \frac{][}{[\quad v_2 \quad]} + \frac{[\quad u_3 \quad]}{]v_3[} *$$

Again, we have also added new information. The first + implies that the starting point of v_1 is after the starting point of u_1. The first $*$ indicates that u_1 and v_2 do not have to overlap. They could be only neighbours. The second +, however, tells us that v_2 and u_3 do properly overlap. The second $*$ informs us, that it is possible that u_3 and v_3 end at the same position.

Step 5. As a last step, we add the overlap words to the operator symbols, but only in between the terms this time.

$$+\frac{[\quad u_1 \quad]}{[v_1]} *s_1 \frac{][}{[\quad v_2 \quad]} +s_2 \frac{[\quad u_3 \quad]}{]v_3[} *$$

Here s_1 might be empty, but s_2 is not empty. We could have introduced the overlap words for the first and the last operator as well, but simply choosed not to do so here. It is often the case that the first and the last overlap word do not have much significance.

In the first example the formula (1) could represent a double-factorization $x_1x_2x_3 = y_1y_2$, with $x_2 = y_1 = y_2 = v$. The set of factor points could be $\{(0,r),(1,l),(1,r),(2,l),(2,r)\}$. The set of constraints could then be $\big\{(1,0,\{\succ,\succeq\}),(1,1,\{\prec,\preceq\}),(2,1,\{\succ,\succeq\}),(2,2,\{\prec,\preceq\})\big\}$.

References

[1] C. Choffrut, J. Karhumäki, *Combinatorics on Words* in: G. Rozenberg, A. Salomaa (Eds.): Handbook of Formal Languages **1**, Chapter 6, 329–438, Springer 1997.

[2] E. Czeizler, Š. Holub, J. Karhumäki, M. Laine, *Intricacies of simple word equations: an example*, International Journal of Foundations of Computer Science, World Scientific, submitted 2007.

[3] S. Holub, *Binary equality sets are generated by two words*, Journal of Algebra, **259**, 1–42, 2003.

[4] M. Lothaire, *Combinatorics on words*, Addison-Wesley, 1983.