数理解析研究所講究録
1489巻 2006年 147-152



# 文字列方程式における反復文字列

植村仁 (真理大學 資訊科學系)

## 1 はじめに

文字列は計算機科学の領域における基本的要素の1つである. 当領域では文字列の探索, 照合, 比較などをしばしば行う必要があるが同一の文字列の接頭語と接尾語が等しい場合, その文字列がより短い文字列の反復となることが散見 [1, 2] される. 本稿では文字列上の方程式という観点からある関係を持つ文字列が反復される文字列をもつことについての諸相を述べる.

## 2 文字列方程式とは

$\Sigma$ を定数の有限集合, $V$ を加算無限な変数の集合とし, $V\cap\Sigma=\emptyset$ とする. 変数を $X, X_1, X_2, \cdots$, $Y, Y_1, Y_2, \cdots$, $Z, Z_1, Z_2, \cdots$ により, 定数を $a, b, c, a_1, a_2, \ldots$ 等で表す. 本稿では文字列は $\Sigma\cup V$ 上の有限文字列であるとする. 定数からなる文字列を特に定数列とよぶ. 文字列 $\alpha$ の 長さを $|\alpha|$ で表す. 長さが 0 の文字列を空列と呼び $\varepsilon$ で表す. $\Sigma^*$ は空列を含む定数列全体, $\Sigma^+$ は空列を含まない定数列全体からなる集合であるとする. また, $T$ を文字列の集合とするとき $T^*$ を $T$ の要素から0回以上任意に選び連接させたもの全体からなる集合, $T^+$ を $T$ の要素から0回以上任意に選び連接させたもの全体からなる集合とする.

文字列方程式とは変数と定数からなる文字列の等式である. 文字列方程式が複数本存在し連立するとき, この系を連立文字列方程式と呼ぶ.

代入とは 置換 $X:=\alpha$ $(X\in V,\ \alpha\in(\Sigma\cup V)^*)$ の集合である. ただし同一変数が2回以上出現しないものとする. 文字列方程式の解とは, 等式を成立させる代入のうち全ての変数を定数列で置き換えるものとする. また連立文字列方程式においてはすべての等式を成立させるものとする. 文字列方程式の解集合とはその解すべてからなる集合であるとする.

## 3 文字列方程式の分類

文字列方程式は以下のように分類することができる.

1. 定数・変数に制限なし, 連立を許す
2. 定数・変数に制限なし, 式の数は1つ
3. 定数なし, 連立を許す
4. 定数なし, 式の数は1つ

定数を含む文字列方程式は変数のみの等式と $X=w$ $(w\in\Sigma^*)$ の形をした等式からなる連立文字列方程式に還元することができる.

本研究ではある変数のみの非連立文字列方程式に着目し諸性質を導く.

## 4 互換非連立文字列方程式

等式のうち, 変数のみからなりそれに出現する変数は両辺に1度ずつであるようなものを互換等式と呼ぶことにする. また, 互換非連立文字列方程式とは互換等式からなる方程式であるとする.

$$例\ X_1X_2X_3X_4 = X_3X_1X_2X_4$$

明らかに, 互換等式の両辺に出現する変数列の長さは等しい.

以下混乱のない限り互換非連立文字列方程式を互換方程式と略する.

## 5 正規互換非連立文字列方程式

互換方程式において両辺に同じ変数の列が出現する場合がある. 文字列方程式の解くことは方程式を成立させる文字列を発見することであるから, 両辺に同じ順序で連続して出現する変数の列を1つの変数に置換しても大きく意味は変わらない. 例えば, $X_1X_2X_3X_4 = X_3X_1X_2X_4$ の $X_1X_2$ を $X$ に置換すると, $XX_3X_4 = X_3XX_4$ となる. つまり

$$\begin{cases} XX_3X_4 = X_3XX_4 \\ X = X_1X_2 \end{cases}$$

と等しい解集合を持つことが分かる. $X = X_1X_2$ は $X_1$ と $X_2$ によりどのように $X$ に代入される定数列を分割するかを表すのみであるので, 本稿ではこのような場合を考察しない.

等式が正規であるとは, 両辺に同順序で連続して出現する変数の列が存在しないこととする. また正規互換非連立文字列方程式とは正規な等式1つからなるものであるとする.

以降誤解のない限り, 正規互換非連立文字列方程式を単に正規互換方程式と記す.

## 6 正規互換非連立文字列方程式の分解

正規互換方程式は解集合が等しく, より次数の低い複数の互換等式からなる連立方程式に分解することができる場合がある.

まず互換等式に対して次数を定義する. 互換等式が $n$ 次であるとはそれが $n$ 種類の変数を含むものであるとする. 互換方程式 (連立もしくは非連立) の次数は, 含まれるすべての等式の次数の最大値であるとする.

**補題 6.1.** $\alpha_i, \beta_i \in V^*$ $(1 \leqq i \leqq n)$ とし, 変数の列 $\alpha \in V^*$ に対して, $var(\alpha)$ をその列に含まれるすべての変数からなる集合であるとする. 正規互換非連立文字列方程式 $\alpha_1 \cdots \alpha_n = \beta_1 \cdots \beta_n$ が

$$\begin{aligned} var(\alpha_i) &= var(\beta_i) \\ var(\alpha_i) &\neq var(\alpha_j)\ (i \neq j,\ 1 \leqq i, j \leqq n) \\ var(\beta_i) &\neq var(\beta_j)\ (i \neq j,\ 1 \leqq i, j \leqq n) \end{aligned}$$

を満たすとき, 以下の正規互換連立文字列方程式は上の方程式と等しい解集合を持ち, $n \geq 2$ のとき, その次数はより低いものとなる.

$$\begin{cases} \alpha_1 = \beta_1 \\ \vdots \\ \alpha_n = \beta_n \end{cases}$$

このような分解ができない正規互換非連立文字列方程式を原子互換方程式と呼ぶことにする.

## 7 反復文字列

次節では正規互換非連立文字列方程式の解には反復される文字列が出現することについて述べる. そのためここでは反復文字列とそれに関する定義を行う.

まず, $u \preceq_p v$ は $u$ が $v$ の接頭語であること, $u \preceq_s v$ は $u$ が $v$ の接尾語であることとする. 接頭語と接尾語の性質から, 次の補題が成立する.

**補題 7.1.** *[2]* $v \in \Sigma^+$, $w \in \Sigma^*$ とする. 以下のことが成立する.

$$(i)\quad \exists i_0 \geq 1\ s.t.\ w \preceq_p v^{i_0}w \iff \forall i \geq 0,\ w \preceq_p v^i w,$$

$$(ii)\quad \exists j_0 \geq 1\ s.t.\ w \preceq_s wv^{j_0} \iff \forall j \geq 0,\ w \preceq_s wv^j.$$

ただし, $v^i$ は定数列 $v$ を $i$ 回反復して連接し得られる定数列であるとする.

定数列 $w$ が反復文字列であるとはある定数列 $u$ と正整数 $i$ が存在して, $w = uu^i$ となることである. また, 文字列 $w$ が準反復文字列とはある文字列 $u, v$ と正整数 $i$ が存在して, $w = (uv)^i u$ となるものである. ただし $u \neq \varepsilon$ とする.

この補題により, 次節からの定理が証明される.

## 8 2次正規互換方程式

1次の互換方程式は自明な解のみを持つので省略する. 2次の互換方程式は変数名を付け替えたものを除き2つだけ存在する. 正規でない互換方程式 $X_1X_2 = X_1X_2$ の解は自明であるので原子互換方程式 $X_1X_2 = X_2X_1$ の解について考察する.

**定理 8.1.** $X_1X_2 = X_2X_1$ の解を $\{X_1 := u_1, X_2 := u_2\}$ とするとある非反復文字列 $u \in \Sigma^+$ が存在して,

$$u_1, u_2 \in \{u\}^*$$

$$u_1u_2 \in \{u\}^*$$

となる.

## 9 3次原子互換方程式の解の性質

**3次互換方程式の分類** 変数名の付け替えたものを除き3次の互換方程式は以下のとおりである.

まず $X_1X_2X_3 = X_1X_2X_3$ は正規ではなく1次方程式と見なすことができ, $X_1X_2X_3 = X_2X_3X_1$, $X_1X_2X_3 = X_3X_1X_2$ もまた正規ではなく2次方程式と見なすことができる. 次に $X_1X_2X_3 = X_1X_3X_2$, $X_1X_2X_3 = X_2X_1X_3$ は1次と2次の等式からなる連立方程式に分解することができる. $X_1X_2X_3 = X_3X_2X_1$ のみが原子互換方程式となることがわかる.

最後の互換方程式のみが原子互換方程式であり, 他のものは正規でないか正規であっても原子互換方程式ではないため, 2次以下の連立した原子互換方程式へと分解される. つまり解に含まれる定数列の性質は2次以下の方程式の解に含まれる定数列の性質を組み合わせたものになる. したがってここでは $X_1X_2X_3 = X_3X_2X_1$ の性質を明らかにする.

**定理 9.1.**

$$X_1X_2X_3 = X_3X_2X_1$$

の解の *1* つを $\{X_1 := u_1, X_2 := u_2, X_3 := u_3\}$ とすると

ある定数列 $u,v$ が存在して, $u_1u_2u_3$ は $u,v$ からなる準反復文字列となり, $u_1,u_2,u_3 \in \{u,v\}^*$ が成立する

## 10 4次以上の正規な長さ限定の正規互換方程式の解の性質

**長さ限定の互換方程式** 4次以上の原子互換方程式の解における反復文字列について議論をするためには, 解に出現する定数列の長さを限定した等式を定義しておく必要がある. また通常の互換方程式と同様に, この制限された方程式に対しても正規なもの, 低次なものに分解されないものを定義する.

$V'$ を長さ限定の変数の集合とする. $V'$ の要素を $X_i[n]$ $(n \in N)$ で表す. 長さ限定の互換方程式は, 変数の集合を $V'$ としてすでに定義した互換方程式と同様に定義される.

長さ限定の互換方程式に対する代入は置換 $X := \alpha$ $(X \in V',\ \alpha \in (\Sigma \cup V)^*)$ の集合によって表現されるが, $c(\alpha)$ を $\alpha$ に出現する定数の数であるとするとき, 各置換 $X_i[n] := \alpha$ は, $\sharp c(\alpha) \square n$ でなければならないものとする. ただし, $\sharp$ は集合の濃度を表す.

長さ限定の互換方程式の解と解集合, 長さ限定の正規互換方程式, 長さ限定の原子互換式についての定義は, すでに定義した互換方程式と同様に定義されるとする.

$$\text{e.g., } X_1[1]X_2[1]X_3[1]X_4[1]X_5[4] = X_5[4]X_4[1]X_3[1]X_2[1]X_1[1]$$

**定理 10.1.** $n \geq 4$ とする. $n$ 次の長さ限定の原子互換方程式の解の *1* つを $\{X_i := u_i \mid i = 1,\cdots,n\}$ とするとある定数列 $u,v$ が存在して,

$u_1 \cdots u_n,\ u_1,\cdots,u_n \in \{u,v\}^*$ が成立する

## 11 その他の注記すべき結果

ここでは互換方程式ではないがその解に出現する定数列に類似した性質を持つ文字列方程式をいくつか取り上げる.

**ある自由変数が存在するものについて**

$$X_1X = XX_2$$

の解の1つを $\{X_1 := u_1, X_2 := u_2, X := v\}$ とするとある $u',u''$ と 正整数 $i,j$ が存在して,

$$v = u_1^i u'' = u'u_2^j \text{ かつ } u_1 = u''u',\ u_2 = u'u''$$

この場合は変数を置換する定数列の長さを限定することなく準反復文字列の性質が出現し, 3次の原子互換式に類似した性質が得られる. その証明自体も類似したものである.

**ある正規でないものについて** 両辺に連続して出現する変数の列を1つの変数とみなせることから正規でないものを除外して議論を進めてきたが, 正規でない非連立互換文字列方程式にも反復文字列について幾つかの性質がある. ここではその1つをとりあげる. 下の $XX_2$ が両辺に連続して出現する変数の列であることに注意する.

$$X_1(XX_2) = (XX_2)X_1$$

の解の1つを $\{X_1 := u_1, X_2 := u_2, X := w\}$ とするとある $u, u'$ $(u \neq \varepsilon)$ と正整数 $i, j$ が存在して, $w = u^i u'$, $u_1(wu_2)w = u^j u'$ となる.

**ある連立した方程式について (1)** 次の結果は両辺に同一変数が出現しないが, 反復される定数列が解に出現するものである. 反復文字列を生じる原因は2つあり, 1つはこの方程式が連立の形であることであり, もう1つは右辺にある同一変数の反復である.

$$\begin{cases} ZY = X_1 \cdots X_1 \\ YZ = X_2 \cdots X_2 \end{cases}$$

の解の1つを $\{X_1 := u_1, X_2 := u_2, Y := v, Z := w\}$ とするとある $u', u'' \in \Sigma^+$ が存在して, $w = u_1 \cdots u_1 u'$ かつ $u_1 = u'u''$ かつ $u_2 = u''u'$ となる. ただし, 各右辺に1つは変数があるものとする.

**ある連立した方程式について (2)** 次の結果は連立して4つの変数が出現し関係するにも拘らず反復する定数列の性質が出現する場合である.

$$\begin{cases} X_1(ZX_2) = (ZX_2)X_1 \\ (X_2Z)X_3 = X_3(X_2Z) \end{cases}$$

の解の1つを $\{X_1 := u_1, X_2 := u_2, X_3 := u_3, Z := w\}$ とすると, ある $\varepsilon$ ではない定数列 $u, u' \in \Sigma^+$ と非負整数 $i, j$ が存在し, $w \preceq_p u^{i+1}$, $u_1 w u_2 w u_3 \preceq_p u^{j+1}$ かつ $w = u^i u'$, $u_1 w u_2 w u_3 = u^j u'$ となる.

**反復文字列と準反復文字列についての結果** 次の場合はある定数列が反復文字列と準反復文字列の性質を持つとき, どのような結果となるかを示す一例である. 1番目の等式は $Z$ に代入される定数列が反復文字列を指し, 2番目の等式は2次の原子方程式と考えることができることから準反復列が変数に代入されることを示している.

$$\begin{cases} Z = X \cdots X \\ X_1 Z = Z X_2 \end{cases}$$

の解の1つを $\{X_1 := u_1, X_2 := u_2, X_3 := u_3, Z := w\}$ とすると, $u$ が空列でなく, $|u_1| \square |w|$ であるとき $u_i \in \{u\}^*$ $(i = 1, 2)$ となる.

## 12 今後の課題

本研究では文字列方程式と反復文字列についてのいくつかの結果を得た．その結果は変数のみからなる文字列方程式には解に反復文字列が生じる必要十分条件が存在することを示唆するがその発見には至らなかった．その他の注記すべき結果にあるように，反復文字列は原子互換方程式にのみ生じるものではなく，考慮される文字列方程式が一般的のものである必要がある．従って一般の文字列方程式についての諸性質を解明することが課題の１つである．

定数を含む文字列方程式の解明も課題の１つである．定数列は長さが決まっているため変数のみからなりかつ，いくつかの変数の解の長さを指定することのできる文字列方程式についての性質の解明は定数を含む文字列方程式の諸性質の解明のための糸口となる．よって第一の課題に加えて，代入される定数列の長さを制限した変数の混在までもを許す文字列方程式の解明が第二の課題となる．

## 参考文献

[1] M. Crochemore and W. Rytter: *Jewels of stringology*, World Scientific Publishing, (2002).

[2] Jin Uemura and Masako Sato: *Learning of Erasing Primitive Formal Systems from Positive Examples*, The Special Issue on Algorithmic Learning Theory for ALT 2003 in Theoretical Computer Science, to appear.