Dynamic Arabic mathematical fonts

M. Banouni, M. Elyaakoubi and A. Lazrek

University Cadi Ayyad, Marrakech, Morocco

http://www.ucam.ac.ma/fssm/rydarab

lazrek@ucam.ac.ma

TUG 2004

Outlines

1 Overview

1.1 The problem

Creating a *complete and homogeneous* mathematical font, for typesetting mathematics in an Arabic presentation, observing the calligraphic rules, is a complex artistic and technical task.

Arabic mathematical expression:

$$\text{د(س)} = \begin{cases} \underset{\text{ب} = 1}{\overset{\text{ص}}{\text{مجـــــــــــ}}} \text{س}^{\text{ب}} & \text{إذا كان س} > 0 \\ \int_{1}^{\text{ص}} \text{س}^{\text{ب}} & \text{إذا كان س} < 0 \\ \text{جا}\ \pi & \text{غير ذلك} \end{cases}$$

the same expression in a Latin based writing:

$$f(x) = \begin{cases} \sum_{i=1}^{s} x^i & \text{if} \quad x < 0 \\ \int_{1}^{s} x^i & \text{if} \quad x > 0 \\ \sin\pi & \text{otherwise} \end{cases}$$

1.2 Arabic mathematical language

Arabic writing

- the direction of writing spreads out from right to left
- the Arabic alphabet is special (ا ب ت)
- the cursivity of the writing (محمد)
- the consonants are marked as letters while vowels are diacritics (مُحَمَّدٌ)
- some letters differ only by the number and the position of dots (ب ت ث)
- some letters differ only by some parts of glyphs (ل ن س ص)
- the shape of a letter depends on its position in the word (initial, median, final and isolated form: م مـ مـ م)

- punctuation marks present particularities of orientation (؟) & glyph (·)
- the letters of a word can be superposed through ligatures as characters (محمد محمد)
- the letters can be stretched in a curvilinear way through the kashida, which is an extensible curve (محـــــمد)
- some letters and words can also be superposed as non characters الرياضيات العربية
- several calligraphic styles can be in use. Shapes of letters, ligatures, calligraphic rules, . . . vary according to these styles (Farisy Koufy Maghriby Naskh Thuluth Rouqaa Dywany)

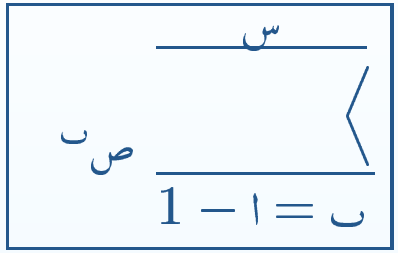
فارسي كوفي مغربي نسخ ثلث الرقعة الديواني ...

Arabic mathematical notation

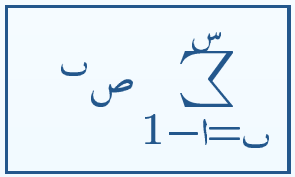
- alphabetic symbols are with or without dots
  (ب as i or ـ as i)
- the alphabetic order in use is the (ا ب ج د) order
  instead of the (ا ب ت ث)
- no cursivity is applied between adjacent alphabetic
  symbols (لاحـ)
- there are two numeral systems (٢١٠ and 210)
- superscript and subscript are on the left ($_2$ـ ^{2}ـ)
- abbreviations use connected letters with ligatures but
  without vowels; with or without dots
  (جا as $\sin$ or حا as sin)

- there are two punctuation systems ($3{,}14$ and $(3{،}\ 14)$ح)
- some symbols are mirrored;

  they can be extensible

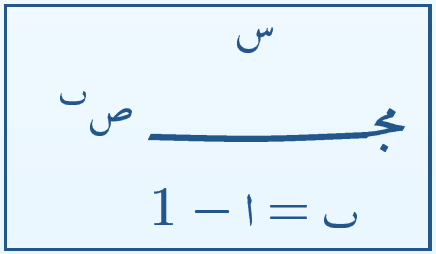
  

  or not

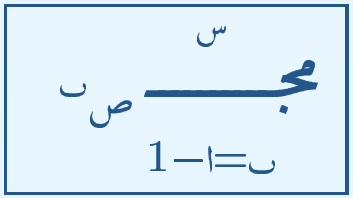
  

- some conventional calligraphic symbols are with

  a curvilinear Kashida

  instead of the linear Kashida

There are no attempts to build a complete and homogeneous font for Arabic mathematics except:

- NasX and Arabic Computer Modern fonts used by the `RyDArab` system for Arabic e-documents composition

- Proofs of some symbols in `TrueType` format by Fayez ALHARGAN

2 RamzArab font

2.1 Elements

RamzArab, font in OpenType format, is composed of:

- alphabetic symbols (ب ـب ـبـ ـبـه) ▷
- punctuation marks (؟ ! : ؛ ،)
- two numeral systems for digits: ▷
  - Arabic: 0 1 2 3 4 5 6 7 8 9
  - Arabic-Indic: ٠١٢٣٤٥٦٧٨٩
- accents, to be combined with alphabetic symbols (◌̇ ◌̈ ◌̃)
- ordinary mathematical symbols such as delimiters, arithmetic operators, ... (( ) [ ] + -)
- mirror image of some symbols (√ ∫ ∑)

2.2 Production

* Composition of specific mathematical Arabic symbols:

- Drawing the whole family of characters by hand (by a calligrapher)
- Scanning the proofs and transforming them into vectors (by tools such as Font Creator or Autotrace) then, achieving their design
- Processing and analyzing the characters then, generating the file defining the font (by software such as Font Creator or PfaEdit)

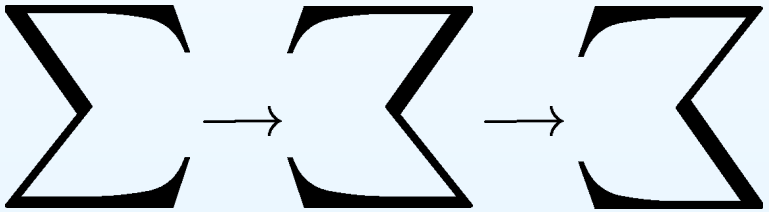
* Composition of ordinary symbols:

- Ordinary symbols can be borrowed from some fonts without any transformation

* Composition of reflected symbols

- Mirrored symbols are taken from the Computer Modern font family. Thus, they are rotated with respect to the vertical axis in the middle of the symbol.

In particular, the sum symbol may be reversed according to the vertical and the horizontal axis in order to follow the orientation of the feather's head in Arabic calligraphy:



3 Dynamic fonts with `CurExt`

3.1 Extensibility

The extensibility is performed according to the parameters:

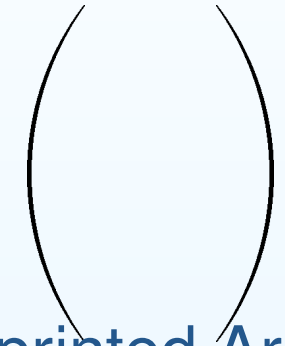
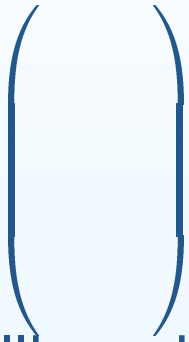
- height: $\int$, $\backslash$, [,], $\Uparrow$
- length: $\underline{abc}, \overline{abc}, \widehat{abc}, \widetilde{abc}, \overleftarrow{abc}, \overrightarrow{abc}, \overbrace{abc}, \underbrace{abc}$

  $\underbrace{\text{ابح}}$ ، $\overbrace{\text{ابح}}$ ، $\overrightarrow{\text{ابح}}$ ، $\overleftarrow{\text{ابح}}$ ، $\widetilde{\text{ابح}}$ ، $\widehat{\text{ابح}}$ ، $\overline{\text{ابح}}$ ، $\underline{\text{ابح}}$
- bidimensionality: $\sqrt{}$, $\sqrt{}$

The rendering of the extensibility can be:

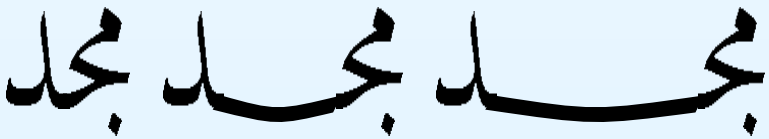
- rectilinear: [,], $\longrightarrow$
- curvilinear: (,), $\int$

Both mathematical expressions and Arabic texts need some curvilinear variable-sized symbols. The curvilinear extensibility may be done in a correct way through producing curves in variable sizes instead of linear segments.

* Parentheses:  instead of 

* Kashida:

In high-quality printed Arabic works, the calligraphy uses the Kashida, a curvilinear variable lengthening of letter along the baseline. The Kashida is also used for the justification of the Arabic text in some words:



and in a mathematical expression in some symbols:  instead of 

3.2 Production of the variable-sized symbols

- *Through measuring*: glyphs sizes are taken from the context of the symbols

* there is one glyph per size through a dynamic font
* a very high precision can be obtained but a second processing is necessary, after a first one where measurements of sizes are taken and recorded

- *Ready to wear*: standardized sizes are determined then, glyphs are drawn according to this fixed set of sizes

* there is one glyph per interval of sizes through static font
* no second processing is needed but the precision is weak

- *Semi-finished*: combination of the two previous ways of doing

3.3 Current extensibility

Extensibility with TEX

TEX system allows curvilinear extensibility through:

- the *a-priori* production of curvilinear glyphs for some sizes: $\left(\left(\left(\left(\left(\quad\right)\right)\right)\right)\right)$
- the composition through curvilinear primitives and linear fragments beyond these sizes: $\left(\quad\right)$

Sometimes, there is no extensibility beyond some size:

$$\widehat{a}, \widehat{ab}, \widehat{abc}, \widehat{abcd}, \widehat{abcde}, \widehat{abcdef}, \widehat{abcdefg}, \widehat{abcdefgh}$$

Other systems allowing extensibility

yhmath : an extension of the number of predefined available sizes
LaTeX package developed by Y. Haralambous

Grif/Thot : allows the production of structured documents with parameterized font `Math-Fly` on PostScript
IRISA and IMAG from Opéra project from Grif SA

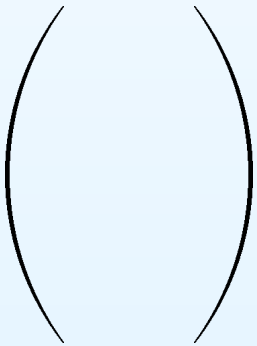
ditroff/ffortid : used for the composition of bidirectional documents with PostScript
Application developed by J. Srouji and D. M. Berry

MathTimeProfessional : Mathematics TeX fonts designed for the Times font family on PostScript
Personal TeX Inc.

3.4 Extensibility with `CurExt`

The `CurExt` system was developed to allow building extensible mathematical symbols in a curvilinear way, using a parameterized dynamic font with the font generator `METAFONT`.

For example, `CurExt` was able to produce automatically a dynamic curvilinear pair of parentheses of any size according to the context



In this adaptation, we propose to use the Adobe PostScript Type 3 format instead of `METAFONT`.

3.4.1 Parameters

The development of a dynamic font requires the determination of some parameters:

- the shape of the glyph according to the dimensions of the character
- the shape of the glyph of small sizes characters and that of big sizes
- the position of the glyph, points of control, according to the dimensions of the character
- the dimensions, width and length, of the box of the character
- the position, height and depth, of the character with respect to the baseline
- the position of the character according to the other characters of the same line, the blank between characters
- the position of the character according to the expression covered by this character

Parenthesis's parameters

Parameters that determine the *parentheses* are:

- the level of concavity
- the shape of the extremities
- the level of thickness of the extremities
- the thickness and the shape of the parenthesis according to the size
- the position of the vertical axis of symmetry to obtain the closing parenthesis starting from the open parenthesis

▷

3.4.2 Production procedure

The production of variable-sized symbols follows the steps:

- collecting the various sizes needed in a parameter file `par`
- generating a file `pl` with the `par2pl` local tool starting from the `par` file according to the POSTSCRIPT syntax
- converting the `pl` file in a metric file `tfm` with the `pltotf` application
- compiling the document to generate a `dvi` file
- generating the `ps` file

This process should be repeated as many times as needed to avoid the overlapping of extensible symbols.

▷

3.4.3 Commands

The syntax of the commands offered by `CurExt` can be as follows:

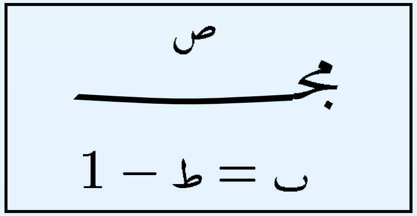
* Parentheses:

```
$\parentheses
 {\matrix{1 & 2 & 3\cr
          4 & 5 & 6\cr
          7 & 8 & 9\cr
 }}$
```

$$\begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{pmatrix}$$

* Kashida:

```
\amarabmath
 ${\csum_{b=T-1}^{s}}$
```



$$\underset{ب = ط - 1}{\overset{ص}{\text{مجـــــ}}}$$

4 Conclusions

The main constraints observed through this work are:

- the observation of the Arabic calligraphy rules, in the Naskh style. These rules are to be formalized
- the necessity of using many digital typography tools, rules and techniques

`RamzArab`, is available as an `OpenType` font, it tries to meet, as far as possible, the requirements of:

- homogeneity: symbols are designed with the same feather. So their shapes, sizes, boldness, and other attributes are homogeneous
- completeness: the set of symbols offered contains most of the usual specific Arabic symbols in use

To preserve some compatibilities, the `RamzArab` font is presented as an extension, not directly used, in TEX environment.

This font is under test for Arabic e-documents after having been structured for the Unicode Standard.

The dynamic component, through `CurExt`, also works in POSTSCRIPT, for parentheses and Kashida. These symbols are representative of the possible cases of horizontal and vertical curvilinear extensibility. So, that can be easily generalized to other variable-sized symbols.

OpenType format

The choice of the OpenType format for the construction of the RamzArab font is due to the following reasons:

- it gives the symbol's outline
- it's a multi-platform fonts
- it become a standard combining the two technologies TrueType and POSTSCRIPT
- it offers advanced typographic layout thanks to its multi-table feature

◁

Alphabetic symbols forms

In the Latin alphabet based writing, characters can be emphasized through **boldness**, use of *italic* or *slanted*, small capitals and so on. In Arabic, there are different ways for varying characters:

- isolated standard form (ب)
- initial standard form (بـ)
- isolated double-struck form (𝕓)
- form with loop (ﺑﻪ)

- initial form with tail (ـہ). The tail is the final form of the letter HEH ہ
- stretched form (ـا). The stretched addition is the final form of the letter ALEF ا

1. The latest two forms are not integrated into the font, because they can be obtained through a simple composition: ہ ← ہ + د ـا ← ا + د
2. In general, literal symbols are without diacritic marks. They can be combined with diacritic marks if necessary.

◁

Confusions

In order to avoid some possible confusions, some non respect of the calligraphic rules would be necessary:

- the letter ALEF ا is denoted by أ, in order to avoid confusion with the digit one ١
- the final form of letter HEH ه is denoted by هـ, in order to avoid confusion with the digit five ٥ and the Roman letter o
- the digit zero ٠ is higher than the full stop dot ٠ both a full lozenge to distinguish them from the Roman dot . that is a full circle
- the isolated letter KAF ك , composed with two elements, will be replaced by the KAF glyph in Ruqaa style ك

◁

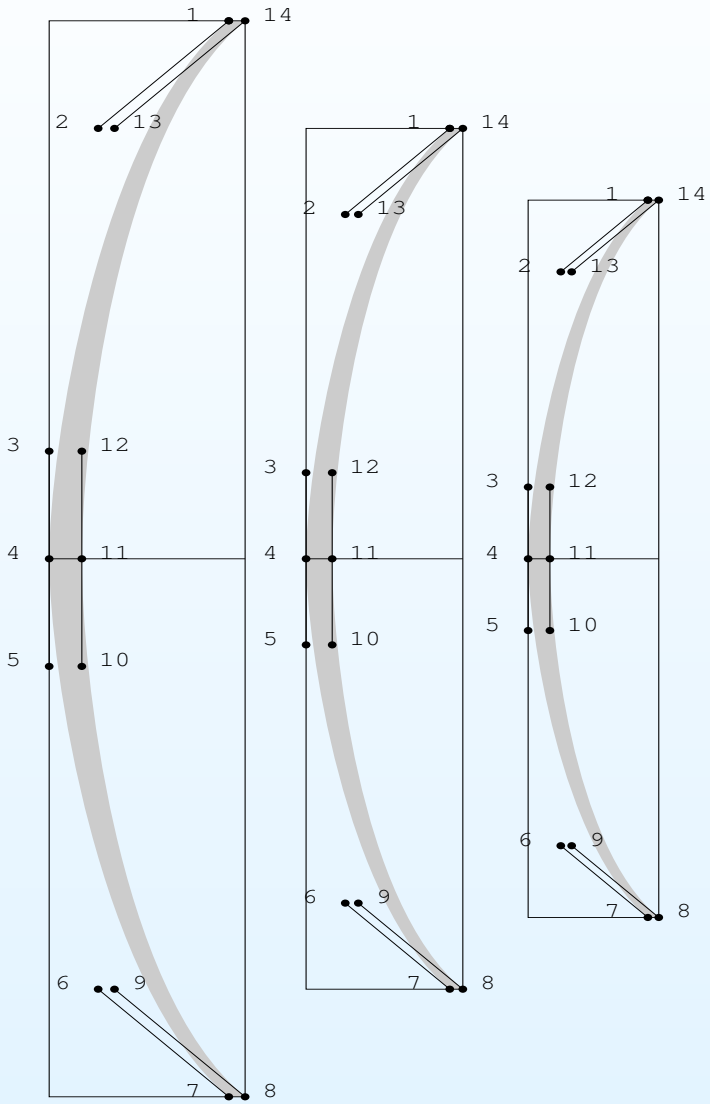
PostScript Type 3

The well known PostScript Type 3 format is characterized by:

- a collection of procedures for drawing individual characters
- an encoding vector for translating each character into the corresponding procedure
- a procedure for calling the correct drawing procedure
- the use of the full PostScript language
- the possibility of programming parameterized procedures

◁

Parenthesis's parameters



Production's process