L2/20-275R

Title: Proposed variation sequences for math calligraphic letters
Author: Murray Sargent
Date: 2021-01-05

> This revised version captures recommendations by the script subcommittee, Asmus Freytag, and John Hudson. The changes include explicit variation sequences and handle capital letters only.

This document explains the need for distinguishing the regular math calligraphic/chancery alphabet from the regular fancy-script/roundhand alphabet. To do so, it lists proposed variation sequences for these alphabets.

In most documents, chancery and roundhand styles can be substituted for one another as a choice of font. Accordingly, when the math alphanumeric symbols were added to Unicode Standard 3.1, the two script styles were unified. But since then, people have documented that the two styles are not always interchangeable.

Chancery and roundhand alphabets are used contrastively by some authors and [La]TeX can support both kinds of letters. Accordingly, mathematicians need a way to distinguish chancery from roundhand in the same document. Variation sequences provide a straightforward way to gain contrastive usage of chancery and roundhand script symbols, especially since some math fonts have chancery letters at the existing math-script code points, while the Unicode Standard has roundhand letters at those code points.

Note that since the lower-case script letters look almost the same in chancery and roundhand, lower-case script letters do not need to be distinguished and we don’t give variation sequences for them. We have not seen contrastive usage of lower-case chancery and roundhand script letters. Nor have we seen contrastive usage of math bold script letters and we don’t give variation sequences for them either. Nevertheless, a font vendor might want to include complete sets of these characters in the event that word processing programs want to display math script letters with uniform styling.

Examples of both chancery and roundhand in the same text

First, here’s an example of script and calligraphic F’s being used in the same document:

> Let $\mathbf{M}^\alpha = (\Omega, \mathscr{F}, \mathscr{F}_t, \theta_t, P_x, X_t)$ be the symmetric α-stable process on $\mathbb{R}^1$ with $0 < \alpha < 2$. Here $\{\mathcal{F}_t\}_{t \geq 0}$ is the minimal (augmented) admissible filtration and θ_t,

And here are examples featuring P’s and C’s in which script letters denote infinity categories

Remark 3.16. Consider a Koszul morphism $\alpha : \Omega\mathcal{C} \to \mathcal{P}$. We have the following diagram at the level of model categories:

$$\begin{tikzcd} \mathcal{P}\text{-}\mathtt{alg} \arrow[r, shift left, "B_\alpha"] \arrow[r, phantom, "\top"] & \mathcal{C}\text{-}\mathtt{coalg}_\alpha \arrow[l, shift left, "\Omega_\alpha"] \arrow[r, shift left, "\mathrm{id}"] \arrow[r, phantom, "\perp"] & \mathcal{C}\text{-}\mathtt{coalg}_\epsilon \arrow[l, shift left, "\mathrm{id}"] \end{tikzcd}$$

There are several roughly equivalent ways to pass from a model category to an infinity category. One is to apply simplicial localization [BK12] followed by the homotopy coherent nerve $\mathcal{N}^{\mathrm{hc}}$ [Lur09]. If we apply this composite functor to our diagram, the leftmost and rightmost entry are well known infinity categories:

$$\begin{tikzcd} \mathfrak{P}\text{-}\mathfrak{alg} \arrow[r, shift left, "B_\alpha"] \arrow[r, phantom, "\cong"] & \mathcal{N}^{\mathrm{hc}}\mathcal{L}\,(\mathcal{C}\text{-}\mathtt{coalg}_\alpha) \arrow[l, shift left, "\Omega_\alpha"] \arrow[r, shift left] \arrow[r, phantom, "\perp"] & \mathfrak{C}\text{-}\mathfrak{coalg}^{\mathrm{nil}}_{\mathrm{dp}}. \arrow[l, shift left] \end{tikzcd}$$

Here we use script lettering to denote infinity categories. We reemphasize on the right that we are working with conilpotent coalgebras and that when the characteristic is nonzero we get the so-called infinity category of divided-power coalgebras.

Here is an example from *Focus on Foci — 0.92*, 22 June 2017 of both styles in print (highlighted):

Remark. If we were to write the inner product in Dirac's bra-ket form as

$$ux + vy + wz = \langle u, v, w | x, y, z \rangle$$

then we could introduce a similar notation for the lines, using, say, doubled up fences $\langle\!\langle$ and $\rangle\!\rangle$ and a separator $\|$, with

$$\begin{aligned} \mathscr{L}_{u,v,w}(x, y, z) &= \langle\!\langle u, v, w \| x, y, z \rangle\!\rangle \\ &= \langle\!\langle u, v, w \| \lambda x, \lambda y, \lambda z \rangle\!\rangle = \langle\!\langle \mu u, \mu v, \mu w \| x, y, z \rangle\!\rangle \end{aligned}$$

for any $\lambda, \mu \in \mathbb{C}^* := \mathbb{C}\backslash\{0\}$; this would emphasize the duality between points and lines. This works only because the condition defining a line $\mathscr{L}_{u,v,w}$ is that $ux + vy + wz = 0$, so multiplication by a non-zero scale factor does not change the condition.

Now we introduce further notation that will lead us to use three different types of non-Roman capital L: script $\mathscr{L}$ for the point sets, calligraphic $\mathcal{L}$ for linear functionals and fraktur $\mathfrak{L}$ for the values, each letter decorated with appropriate argument symbols. We can look at a more general function in the condition above, writing for the moment,

$$\Phi(u, v) := \sum_{k=1}^{n} \prod_{j \in [1,k-1] \cup [k+1,n]} \mathcal{L}_{u,v,-1}(x_j, y_j) = \sum_{k=1}^{n} \prod_{j \in [1,k-1] \cup [k+1,n]} \mathcal{L}_j(u, v) = \sum_{k=1}^{n} \mathfrak{L}_1 \ldots \hat{\mathfrak{L}}_k \ldots \mathfrak{L}_n$$

In addition, both script styles are in the OMS encoding for LaTeX

```
\documentclass{article}
\usepackage{calrsfs}
\DeclareMathAlphabet{\pazocal}{OMS}{zplm}{m}{n}
\newcommand{\La}{\mathcal{L}}
\newcommand{\Lb}{\pazocal{L}}
\begin{document}
$\La\Lb$
\end{document}
```

This LaTeX snippet displays a roundhand L followed by a chancery L

$$\mathscr{L}\ \mathcal{L}$$

Accordingly, although not commonplace, the need for both chancery and roundhand alphabets is attested.

Cambria Math/Unicode quandary

Complicating the addition of new alphabets is the fact that the current math-script alphabets may be chancery in one font and roundhand in another. [Cambria Math](), the first widely used Unicode math font, has chancery letters at the math-script code points, while the Unicode Standard has roundhand letters at those code points. For example, here is the upper-case math-script H (U+210B) in Cambria Math followed by the one in the Unicode Standard:

$$\mathcal{H}\ \mathscr{H}$$

The STIX math fonts had roundhand letters at the math-script codepoints, but in the STIX Two Math font, they have been changed to chancery. This removes the worst conflict in defining the new alphabets, although other math fonts might have roundhand letters at the current math-script codepoints. For more examples, see L2/20-281.

Proposed variation sequences

Define a set of standardized variation sequences in which an uppercase character in the current math-script alphabets is followed by one of two variation selectors U+FE00 and U+FE01 to select a chancery and roundhand style, respectively. If no variation selector is used, display the character as given in the current font. This mirrors the way we use variation selectors (U+FE0E, U+FE0F) for emoji to force text and emoji glyphs, respectively. The variation sequences are listed in Appendix A.

The variation selector approach has the advantages

a) Contemporary software supports variation selectors for East Asia and emoji, so adding new variation selector usage should not be difficult
b) The variation selector U+FE00 is already used with a number of slashed math operators
c) No new code points need to be allocated
d) Most documents can continue to do what they have been doing: ignore the distinction
e) If a math font does not support the variation selectors, it falls back naturally to the current script letters instead of displaying the missing-glyph box (but the style difference is lost)
f) There are eight upper-case script letters in the Letterlike Symbols block (212C $\mathcal{B}$, 2130 $\mathcal{E}$, 2131 $\mathcal{F}$, 210B $\mathcal{H}$, 2110 $\mathcal{I}$, 2112 $\mathcal{L}$, 2133 $\mathcal{M}$, and 211B $\mathcal{R}$). These should not be changed since they were added in Unicode 1.1 from an earlier standard. In particular, $\mathcal{M}$ can be used for the German Mark currency symbol before WWII and when so used, the range of glyph variation is restricted.

Adding two variation selectors for the math script letters may make people ask why we didn't use variation selectors for the math alphabets in the first place, but we all know the arguments in favor of what we did (see the blog post on Math Font Binding). Adding two variation selectors solves the chancery/roundhand quandary quite well, although the use of variation selectors is generally a poor one for situations where symbol shapes need to be used in a contrastive manner—this case should therefore not serve as a general precedent, but should be seen as an exception, tailored to fit this specific case.

Acknowledgements

Thanks to Asmus Freytag, John Hudson, Rick McGowan, Debbie Anderson, Roozbeh Pournader, and Ken Whistler for enlightening discussions that improved this proposal.

Appendix A. Variation sequences for math capital chancery and roundhand letters

The Unicode Standard has mathematical script capital letters in the range U+1D49C..U+1D4B5 with eight characters in the Letterlike Symbols block. The following variation sequences encode the mathematical chancery capital letters

1D49C FE00	𝒜	MATHEMATICAL CHANCERY CAPITAL A
212C FE00	ℬ	MATHEMATICAL CHANCERY CAPITAL B
1D49E FE00	𝒞	MATHEMATICAL CHANCERY CAPITAL C
1D49F FE00	𝒟	MATHEMATICAL CHANCERY CAPITAL D
2130 FE00	ℰ	MATHEMATICAL CHANCERY CAPITAL E
2131 FE00	ℱ	MATHEMATICAL CHANCERY CAPITAL F
1D4A2 FE00	𝒢	MATHEMATICAL CHANCERY CAPITAL G
210B FE00	ℋ	MATHEMATICAL CHANCERY CAPITAL H
2110 FE00	ℐ	MATHEMATICAL CHANCERY CAPITAL I
1D4A5 FE00	𝒥	MATHEMATICAL CHANCERY CAPITAL J
1D4A6 FE00	𝒦	MATHEMATICAL CHANCERY CAPITAL K
2112 FE00	ℒ	MATHEMATICAL CHANCERY CAPITAL L
2133 FE00	ℳ	MATHEMATICAL CHANCERY CAPITAL M
1D4A9 FE00	𝒩	MATHEMATICAL CHANCERY CAPITAL N
1D4AA FE00	𝒪	MATHEMATICAL CHANCERY CAPITAL O
1D4AB FE00	𝒫	MATHEMATICAL CHANCERY CAPITAL P
1D4AC FE00	𝒬	MATHEMATICAL CHANCERY CAPITAL Q
211B FE00	ℛ	MATHEMATICAL CHANCERY CAPITAL R
1D4AE FE00	𝒮	MATHEMATICAL CHANCERY CAPITAL S
1D4AF FE00	𝒯	MATHEMATICAL CHANCERY CAPITAL T
1D4B0 FE00	𝒰	MATHEMATICAL CHANCERY CAPITAL U
1D4B1 FE00	𝒱	MATHEMATICAL CHANCERY CAPITAL V
1D4B2 FE00	𝒲	MATHEMATICAL CHANCERY CAPITAL W
1D4B3 FE00	𝒳	MATHEMATICAL CHANCERY CAPITAL X
1D4B4 FE00	𝒴	MATHEMATICAL CHANCERY CAPITAL Y
1D4B5 FE00	𝒵	MATHEMATICAL CHANCERY CAPITAL Z

The following variation sequences encode the mathematical roundhand capital letters. Note: the roundhand letter glyphs here are accessed by the corresponding ASCII letters with a roundhand font since I don't have an updated Cambria Math font with the roundhand letters.

1D49C FE01	𝒜	MATHEMATICAL ROUNDHAND CAPITAL A
212C FE01	ℬ	MATHEMATICAL ROUNDHAND CAPITAL B
1D49E FE01	𝒞	MATHEMATICAL ROUNDHAND CAPITAL C
1D49F FE01	𝒟	MATHEMATICAL ROUNDHAND CAPITAL D
2130 FE01	ℰ	MATHEMATICAL ROUNDHAND CAPITAL E
2131 FE01	ℱ	MATHEMATICAL ROUNDHAND CAPITAL F
1D4A2 FE01	𝒢	MATHEMATICAL ROUNDHAND CAPITAL G
210B FE01	ℋ	MATHEMATICAL ROUNDHAND CAPITAL H
2110 FE01	ℐ	MATHEMATICAL ROUNDHAND CAPITAL I

1D4A5 FE01	𝒥	MATHEMATICAL ROUNDHAND CAPITAL J
1D4A6 FE01	𝒦	MATHEMATICAL ROUNDHAND CAPITAL K
2112 FE01	ℒ	MATHEMATICAL ROUNDHAND CAPITAL L
2133 FE01	ℳ	MATHEMATICAL ROUNDHAND CAPITAL M
1D4A9 FE01	𝒩	MATHEMATICAL ROUNDHAND CAPITAL N
1D4AA FE01	𝒪	MATHEMATICAL ROUNDHAND CAPITAL O
1D4AB FE01	𝒫	MATHEMATICAL ROUNDHAND CAPITAL P
1D4AC FE01	𝒬	MATHEMATICAL ROUNDHAND CAPITAL Q
211B FE01	ℛ	MATHEMATICAL ROUNDHAND CAPITAL R
1D4AE FE01	𝒮	MATHEMATICAL ROUNDHAND CAPITAL S
1D4AF FE01	𝒯	MATHEMATICAL ROUNDHAND CAPITAL T
1D4B0 FE01	𝒰	MATHEMATICAL ROUNDHAND CAPITAL U
1D4B1 FE01	𝒱	MATHEMATICAL ROUNDHAND CAPITAL V
1D4B2 FE01	𝒲	MATHEMATICAL ROUNDHAND CAPITAL W
1D4B3 FE01	𝒳	MATHEMATICAL ROUNDHAND CAPITAL X
1D4B4 FE01	𝒴	MATHEMATICAL ROUNDHAND CAPITAL Y
1D4B5 FE01	𝒵	MATHEMATICAL ROUNDHAND CAPITAL Z