1 Formal Language Theoretic Definition of NREM Periods

We give a rigorous formulation of the notion of an NREM period using formal language theory. The goal is to encode sequences of sleep stages as words over a finite alphabet and define, precisely and unambiguously, which substrings correspond to valid NREM periods.

1.1 Alphabet and Basic Definitions

Let the finite alphabet
$$\Sigma = \{1, 2, 3, 4, 5, 6\}$$
represent sleep stages, where:

- $\mathcal{N} = \{2, 3, 4\}$ denotes NREM stages,
- $R = \{5\}$ denotes REM sleep,
- $W = \{6\}$ denotes wakefulness,
- $O = \Sigma \setminus \mathcal{N}$ denotes non–NREM stages.

A full night of sleep is represented by a word

$$x = x_1 x_2 \dots x_n \in \Sigma^*.$$

We assume each symbol represents one epoch of fixed duration (e.g. 30 seconds). Let

$$m \in \mathbb{N} \quad \text{and} \quad n \in \mathbb{N}$$

denote thresholds corresponding to:

- m: the minimum number of consecutive epochs defining a terminating sequence (e.g. $m = 10$ epochs for 5 minutes),
- n: the minimum total number of NREM epochs required for a valid NREM period (e.g. $n = 30$ epochs for 15 minutes).

1.2 Runs and Substrings

Given a word $x \in \Sigma^*$, a substring

$$y = x_i x_{i+1} \dots x_j$$

is called a *run* of a set $A \subseteq \Sigma$ if $y \in A^+$ and either:

- $i = 1$ or $x_{i-1} \notin A$, and
- $j = |x|$ or $x_{j+1} \notin A$.

Thus a run is a maximal contiguous block of symbols from A.

1.3 Terminating Sequences

A substring t is called a *terminating sequence* if either

$$t \in R^m R^* \qquad \text{or} \qquad t \in W^m W^*.$$

In other words, t is a maximal run of REM or wakefulness containing at least m consecutive symbols.

1.4 Admissible Interruptions

Let

$$S = O^{<m}$$

denote the set of non–NREM substrings whose length is strictly less than m. These represent interruptions that do not terminate an NREM period.

Intuitively, interruptions shorter than m epochs are ignored for the purposes of continuity.

1.5 Definition of an NREM Period

Let $y \in \Sigma^*$ be a substring. We say that y is a valid *NREM period* if and only if all of the following conditions hold:

1. **Structure with admissible interruptions.**

   The substring y can be written as

$$y = z\,t,$$

where

$$z \in (S^* \mathcal{N})^+$$

and t is a terminating sequence.

Equivalently, z consists of NREM stages possibly interspersed with interruptions shorter than m epochs.

2. **No internal terminating sequences.**

   Within z there is no substring belonging to $R^m R^*$ or $W^m W^*$. That is, no internal run of REM or wakefulness reaches the terminating threshold.

3. **Minimum total NREM duration.**

   Let

   $$|z|_{\mathcal{N}}$$

   denote the number of symbols of z belonging to $\mathcal{N}$. Then

   $$|z|_{\mathcal{N}} \geq n.$$

4. **Termination.**

   The suffix t is a terminating sequence, meaning that the NREM period ends upon the first occurrence of at least m consecutive REM epochs or at least m consecutive wake epochs.

1.6 Intuition

This definition formalizes the following clinical criteria:

- NREM sleep need not be continuous; short interruptions are allowed.
- Interruptions shorter than m epochs are ignored.
- Long runs of REM or wakefulness terminate the period.
- A valid period must contain at least n epochs of NREM sleep in total.

1.7 Language-Theoretic Characterization

Let $\mathcal{T}$ denote the set of terminating sequences. Define the language

$$\mathcal{L}_{\mathrm{NREM}} = \left\{ y \in \Sigma^* \;\middle|\; y = zt,\; z \in (S^*\mathcal{N})^+,\; |z|_{\mathcal{N}} \geq n,\; t \in \mathcal{T} \right\}.$$

Then $\mathcal{L}_{\mathrm{NREM}}$ is a regular language. Indeed, recognition requires only:

- counting consecutive runs up to length m,
- counting total NREM symbols up to threshold n,
- finite memory of the current run type.

Hence there exists a deterministic finite automaton (DFA) recognizing $\mathcal{L}_{\mathrm{NREM}}$.