Pattern Matching and Parameter Identification for Parametric Timed Regular Expressions

Akshay Mambakam, Eugène Asarin, Nicolas Basset, Thao Dang
and not Alexandre Donze

May 11th, 2023

Introduction

- Monitoring, runtime verification, trace analysis for CPS
  - Given a trace of a system detect special fragments (dangerous manoeuvres, arrhythmia, circadian rhythms, a melody).
  - On real or simulated system;
  - On-line or off-line;
  - Searching "Timed texts" (music, speech, Morse, subtitles)
- This work: Parametric Timed Regular Expressions, next best thing for complex timed pattern specification for CPS after STL

Outline

1. Pattern matching and Timed Regular Expressions
2. Parametric Signal Regular Expressions
   - PSRE Intersection Example (Introductory)
   - ECG example
   - Interpreting the Parametric Match-set
   - Formal Description of PSRE
3. Parametric Timed Regular Expressions (Event-Based)
4. Parametric Event-Bounded Timed Regular Expressions
5. Parametric Identification
6. Booleanization and matching
7. Future Work

What is pattern matching?

Problem (Pattern matching for regular expressions)

Given word $w \in \Sigma^*$ and regular expression φ find first/last/all subword(s) v of w that matches φ.

Classical solutions

Algorithms: Often based on automata

Tools: `grep`; RE engines of Perl, Python, etc.

Timed Pattern Matching

Same question for timed signals/all matches

Timed Pattern Matching [UFAM14]

- Syntax of **Timed Regular Expressions** (TRE) [ACM02]:

$$\varphi := \epsilon \mid p \mid \overline{p} \mid \varphi \cdot \varphi \mid \varphi \vee \varphi \mid \varphi \wedge \varphi \mid \varphi^* \mid \langle \varphi \rangle_I$$

  (p propositional variable; I integer interval)

- Semantics defined by match set:

$$\mathcal{M}(\varphi, w) = \{(t, t') \in \mathbb{T} \times \mathbb{T} \mid w \text{ matches } \varphi \text{ on } (t, t')\}$$

- **Timed-Pattern Matching** problem:
  Given a signal and an expression compute the match-set.

Main result of Timed Pattern Matching

Theorem

$\mathcal{M}(\varphi, w)$ is a finite union of 2d zones.

Proof

Structural induction over φ.

A 2d zone is bounded by vertical, horizontal and diagonal lines. Using constraints it can be written as follows:

$$\begin{pmatrix} t \prec c_1 \\ t' \prec c_2 \\ t' - t \prec c_3 \\ c_4 \prec t' - t \\ c_5 \prec t' \\ c_6 \prec t \end{pmatrix}$$

Motivating parameter introductions

With TREs it is possible to specify that

- within 23 seconds after braking the motor stops
- after 30 minutes of overheating (that happens if $T > 130$) a pipe explodes (event E)

Parametric formalisms are more expressive and flexible.
With parameters one can describe a generic pattern: after θ minutes of $(T > \sigma)$ the event E occurs.

Three types of parametric TRE

- Parametric Signal Regular Expressions
- Parametric Timed Regular Expressions (event-based)
- Parametric Event-bounded Timed Regular Expressions

PSRE Intersection Example (Introductory)

PSRE:

$$\phi_1 := (\langle p \rangle_{\theta_1} \cdot \neg p) \land (\neg q \cdot \langle q \rangle_{\theta_2}) \land (\mathtt{true} \cdot \langle p \land q \rangle_{\theta_3} \cdot \mathtt{true}) \tag{1}$$

p is a Boolean (square wave) of period 14.
q is obtained by shifting p by 3 time units.
A Polytope in parametric match-set: $(t' - \theta_2 = 3) \land (t + \theta_1 = 7) \land (\theta_3 \leq 4) \land (t' \leq 10) \land (t \leq 3) \land (\theta_3 \geq 1) \land (t \geq 0) \land (t' \geq 7)$.

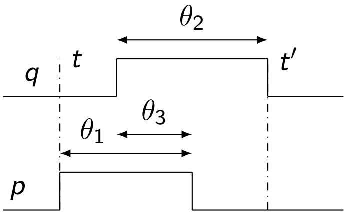


Figure 1: Matching a Parametric Timed Regular Expression ϕ_1

ECG example

PSRE:

$$\phi_2 := \langle -0.55 \le x \le 0.29 \rangle_{\theta_1} \cdot \langle 0.29 \le x \le 2.0 \rangle_{\theta_2} \cdot \langle -0.6 \le x \le 0.29 \rangle_{\theta_3} \quad (2)$$

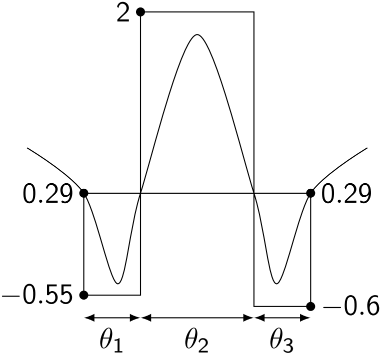


Figure 2: ECG matching expression ϕ_2

Interpreting the Parametric Match-set for ECG Example

- A parametric zone in parametric match-set: $(t + \theta_1 = 227) \wedge (\theta_2 = 6) \wedge (t' - \theta_3 = 233) \wedge (220 \leq t \leq 226) \wedge (234 \leq t' \leq 283)$
- Projecting it on to the parameter space gives the following rectangle: $(1 \leq \theta_1 \leq 7) \wedge (\theta_2 = 6) \wedge (1 \leq \theta_3 \leq 50)$.

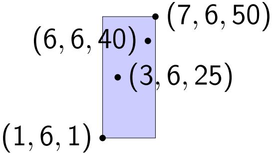
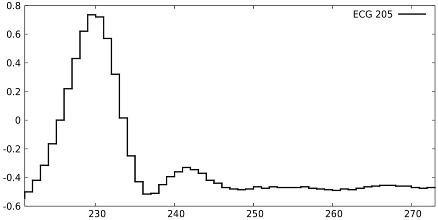
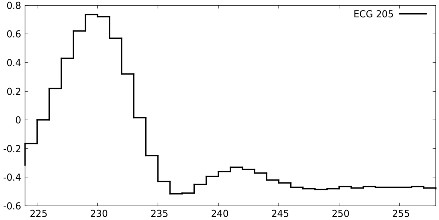


Figure 3: Parameter set (rectangle) from matching

Visualizing the Match



(a) $(t, t') = (221, 273)$ for $(\theta_1, \theta_2, \theta_3) = (6, 6, 40)$



(b) $(t, t') = (224, 258)$ for $(\theta_1, \theta_2, \theta_3) = (3, 6, 25)$

Figure 4: Matching ECG signal (*wecg*)

Parametric Signal Regular Expressions (PSRE)

- Parametric Signal Regular Expressions:

$$p, \overline{p} \mid \langle\phi\rangle_I \mid \phi \cdot \psi \mid \phi^* \mid \phi \wedge \psi \mid \phi \vee \psi$$

$$p := \; x \geq \lambda \mid x \leq \lambda \mid x \geq c \mid x \leq c$$

$I = [a, b]$, where a, b are linear expressions over parameters $(\overline{\theta})$.
$\lambda \in \overline{\theta}$ is a parameter and c is a constant.

Parametric match set:

$$\mathcal{M}(w, \phi(\overline{\theta})) = \{(t, t', \overline{\theta}) \mid (w, t, t') \models \phi(\overline{\theta})\}.$$

Main result of Parametric Timed Pattern Matching

Theorem

The **parametric** *match-set* $\mathcal{M}(\varphi, w)$ *is a finite union of* **parametric** *zones. It is computable knowing expression* φ *and signal* w.

Definition

A parametric zone: $z = \left(\begin{array}{c} t \prec \min(a_1, ..., a_{n_1}) \\ t' \prec \min(b_1, ..., b_{n_2}) \\ t' - t \prec \min(c_1, ..., c_{n_3}) \\ \max(d_1, ..., d_{n_4}) \prec t' - t \\ \max(e_1, ..., e_{n_5}) \prec t' \\ \max(f_1, ..., f_{n_6}) \prec t \end{array} \right) \wedge \mathrm{C_z}(\overline{\theta})$ where

$\mathrm{C_z}(\overline{\theta})$ is a polytope over $\overline{\theta}$.

Method for proof and algorithm

Structural induction over φ.

Parametric zone intersection operation

$$\mathbf{z}_1 = \begin{pmatrix} t \prec c_1 \\ t' \prec c_2 \\ t' - t \prec c_3 \\ c_4 \prec t' - t \\ c_5 \prec t' \\ c_6 \prec t \end{pmatrix} \wedge \boldsymbol{C_{z_1}} \text{ and } \mathbf{z}_2 = \begin{pmatrix} t \prec c'_1 \\ t' \prec c'_2 \\ t' - t \prec c'_3 \\ c'_4 \prec t' - t \\ c'_5 \prec t' \\ c'_6 \prec t \end{pmatrix} \wedge \boldsymbol{C_{z_2}}.$$

Here, c_1, c_2, c_3 can be expressed as minima of linear functions and c_4, c_5, c_6 as maxima of linear functions. Similarly for $c'_1, \ldots, c'_6$. The symbol $\prec$ is $\leq$ / $<$. For intersection of $\mathbf{z}_1$ and $\mathbf{z}_2$:

$$\mathbf{z}_1 \cap \mathbf{z}_2 = \begin{pmatrix} t \prec \min(c_1, c'_1) \\ t' \prec \min(c_2, c'_2) \\ t' - t \prec \min(c_3, c'_3) \\ \max(c_4, c'_4) \prec t' - t \\ \max(c_5, c'_5) \prec t' \\ \max(c_6, c'_6) \prec t \end{pmatrix} \wedge \boldsymbol{C_{z_1}} \wedge \boldsymbol{C_{z_2}}$$

Parametric zone concatenation operation

For concatenation of PTRE and sequential composition of zones z_1 and z_2:

$$\begin{pmatrix} t \prec \min(c_1, c_1' - c_4, c_2 - c_4) \\ t' \prec \min(c_2', c_2 + c_3', c_1' + c_3') \\ t' - t \prec c_3 + c_3' \\ c_4 + c_4' \prec t' - t \\ \max(c_5', c_5 + c_4', c_6' + c_4') \prec t' \\ \max(c_6, c_6' - c_3, c_5 - c_3)) \prec t \end{pmatrix} \wedge C_{z_1} \wedge C_{z_2}$$

$$\wedge (c_5 \prec c_1') \wedge (c_6' \prec c_2) \wedge (c_4 \prec c_3) \wedge (c_4' \prec c_3') \wedge (c_5 \prec c_2) \wedge (c_6' \prec c_1')$$

Idea for algorithms for intersection and concatenation.

- Handle parametric zones by ordering them along the time dimensions.
- Zones well separated in the time dimensions are not related i.e.
- empty intersection and sequential composition.
- Exploit this and avoid useless operations. Called plane-sweep technique.

PTRE (Event-Based) Syntax

Syntax:

$$\underline{a} \mid \langle\varphi\rangle_I \mid \varphi_1 \cdot \varphi_2 \mid \varphi_1 \vee \varphi_2 \mid \varphi_1 \wedge \varphi_2 \mid \varphi^*$$

where $I = [\alpha, \beta]$ and each of α, β is either a non-negative constant or a timing parameter s_i. For all $a \in \Sigma$ we define $\underline{a}$ which represents an arbitrary passage of time followed by event a where Σ is the event alphabet over which the PTRE is defined.

PTRE (Event-Based) Time-Event Sequence Semantics

Let us consider the following PTRE which has both concatenation and intersection operators:

$$\phi_3 := (\langle \underline{a} \cdot \underline{b} \rangle_{\theta_1} \cdot \underline{c}) \wedge (\underline{a} \cdot \langle \underline{b} \cdot \underline{c} \rangle_{\theta_2})$$

Recall that $\underline{a}$ represents an arbitrary passage of time followed by the event a. This can be represented as $r \cdot a$, where r represents passage of time and a is an event.

The semantics of the expression containing parameters θ_1 and θ_2 can be deduced as follows:

$$\{r_1 \cdot a \cdot r_2 \cdot b \cdot r_3 \cdot c : (r_1 + r_2 = \theta_1) \wedge (r_2 + r_3 = \theta_2)\}.$$

PTRE (Event-Based) Match-Set Semantics

The parametric match-set of a PTRE expression φ (with event-based semantics) for a timed word $\omega = t_1 a_1 ... t_n a_n$ is defined inductively as follows ($t_0 = 0$ by default):

$$\mathcal{M}(\underline{a}, \omega) := \{(t, t', v) : \exists i \in [1...n] \cdot a = a_i \wedge t = t_{i-1} \wedge t' = t_i\}$$

$$\mathcal{M}(\langle\varphi\rangle_I, \omega) := \{(t, t', v) : t' - t \in I_v \wedge (t, t', v) \in \mathcal{M}(\varphi, \omega)\}$$

$$\mathcal{M}(\varphi \cdot \psi, \omega) := \{(t, t', v) : \exists t''.(t, t'', v) \in \mathcal{M}(\varphi, \omega) \wedge (t'', t', v) \in \mathcal{M}(\psi, \omega)\}$$

$$\mathcal{M}(\varphi \wedge \psi, \omega) := \{(t, t', v) : (t, t', v) \in \mathcal{M}(\varphi, \omega) \wedge (t, t', v) \in \mathcal{M}(\psi, \omega)\}$$

$$\mathcal{M}(\varphi \vee \psi, \omega) := \{(t, t', v) : (t, t', v) \in \mathcal{M}(\varphi, \omega) \vee (t, t', v) \in \mathcal{M}(\psi, \omega)\}$$

$$\mathcal{M}(\varphi^*, \omega) := \{\exists k \geq 0.(t, t', v) \in \mathcal{M}(\varphi^k, \omega)\}$$

Match-Set and Parametric Intervals

Definition

Parametric Intervals Parametric intervals are of the form:

$$\mathbf{y} := (t = d_1 \wedge t' = d_2 \wedge \boldsymbol{C_y})$$

where d_1, d_2 are real constants and $\boldsymbol{C_y}$ is a polytope over the parameter space.

Theorem

For a given timed word the parametric match-set of a PTRE is a finite union of parametric intervals.

Idea for algorithms for intersection and concatenation.

Sort according to the start and end of intervals. Then use binary search.

Parametric Event-Bounded Timed Regular Expressions

- We can combine the semantics PTRE (event-based) and PSRE (state-based) like [FMNU15].
- We can use switching of value of Boolean signals to indicate events.
- Rising edge of Boolean signal p is indicated $\uparrow p$ which captures change from false to true.
- Note that falling edge can be defined as $\downarrow p := \uparrow \overline{p}$.
- Parametric E-TRE (PE-TRE) syntax: $\uparrow p$, $\psi_1 \cdot \varphi \cdot \psi_2$, $\psi_1 \vee \psi_2$, or $\psi_1 \wedge \varphi$.
- Here, ψ_1 and ψ_2 are PE-TRE and φ stands for a PSRE.

Theorem

For Boolean, piecewise linear and piecewise constant signals the parametric match-set of PE-TRE is always a finite union of parametric intervals.

PE-TRE Example

PE-TRE ϕ_4 denotes the pattern of a brake control signal b for a vehicle under heavy braking situation.

$$\phi_4 := \uparrow b \cdot \langle b \rangle_{[0,\theta_1]} \cdot {\langle \neg b \cdot b \rangle^+}_{[0,\theta_2]} \cdot \downarrow b \tag{3}$$

It starts with a rise edge of b followed by a braking period of duration less than θ_1. It continues with one or more pulses with duration less than θ_2. It ends with a falling edge of b.

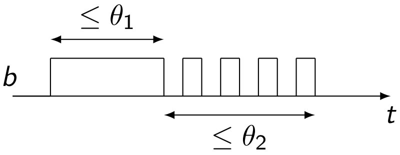


Figure 5: Braking Pattern

Formally Defining Parametric Identification

- Given a PSRE ψ (defined over parameter vectors u and v)
- Signal w
- A list $\mathcal{I}$ of n intervals $I_1 = [a_1, b_1], \ldots, I_n = [a_n, b_n]$.
- Compute the set of parameters that produce a match at each of these n intervals.

Solution Set

- Compute $\mathcal{P}(\psi, w, \mathcal{I}) :=$
- $\{(u, v) : \bigwedge_{1 \leq k \leq n} (\exists t, t' \; (t, t', u, v) \in \mathcal{M}(\psi, w) \wedge t = a_k \wedge t' = b_k)\}$

Parametric identification to identify ECG

PSRE for characterizing and identifying ECG:

$$\phi_7 := \langle -q_0 \le x \le q_1 \rangle_{[0,q_2]} \cdot \langle q_1 \le x \le q_0 \rangle_{[0,q_3]} \cdot \langle -q_0 \le x \le q_1 \rangle_{[0,q_2]}$$

- We label the first ten pulses of ECG using ten intervals of 80 time units each.
- Parametric identification takes 478s. Inefficient for large data.

How to side-step this inefficiency?

- Bottleneck is computing match-set when magnitude parameters (λ) for expressions like $x \le \lambda$.
- It contains number of polytopes equal to the size of the signal.
- Avoid magnitude parameters. Do this by pre-processing signals using Booleanizers.
- Then detect patterns using expressions with only timing parameters.

Min/Max Operators for Booleanization

- Approximate local minima: $(b_{max} := \max_{[-150,150]} x - x \leq 0.05)$
- Approximate local maxima: $(b_{min} := x - \min_{[-10,10]} x \leq 0.05)$
- Express ECG pulse as a minimum followed by maximum followed by another minimum.
- $\phi_{10} := b_{min} \cdot \langle \texttt{true} \rangle_{[0,\theta_1]} \cdot b_{max} \cdot \langle \texttt{true} \rangle_{[0,\theta_2]} \cdot b_{min}$ where $\theta_1, \theta_2 \in [0, 20]$
- The ECG has around a couple of thousand pulses.
- We get 8726 polytopes computed within 33s.
- Fast. Compared to the parametric identification discussed just before.

Future Work

- Atomic predicates involving integrals of the form $\int_t^{t'} x.dt \leq c$.
- How inserting timing parameters like δ in expression like $\langle \varphi \rangle_{[2-\delta,3]}$ is related to measuring timed robustness.
- Restrict the scope to expressions with only magnitude parameters and efficiently handle them using a data structure that combines 2d zones and boxes.
- Parametric identification of PSRE with monotonicity property can be efficiently performed using queries as it has been done for PSTL.