AMONG Implied Constraints for Two Families of Time-Series Constraints

Ekaterina Arafailova, Nicolas Beldiceanu, Helmut Simonis

AMONG Implied Constraints for Two Families of Time-Series Constraints *

Ekaterina Arafailova[1], Nicolas Beldiceanu[1], and Helmut Simonis[2]

[1] TASC (LS2N), IMT Atlantique, FR – 44307 Nantes, France
{Ekaterina.Arafailova,Nicolas.Beldiceanu}@imt-atlantique.fr
[2] Insight Centre for Data Analytics, University College Cork, Ireland
Helmut.Simonis@insight-centre.org

Abstract. We consider, for an integer time series, two families of constraints restricting the *max, and the sum, respectively, of the surfaces* of the elements of the sub-series corresponding to occurrences of some pattern. In recent work these families were identified as the most difficult to solve compared to all other time-series constraints. For all patterns of the time-series constraints catalogue, we provide a *unique per family parameterised* AMONG implied constraint that can be imposed on any prefix/suffix of a time-series. Experiments show that it reduces both the *number of backtracks*/*time spent* by up to 4/3 orders of magnitude.

1 Introduction

Going back to the work of Schützenberger [20], *regular cost functions* are quantitative extensions of regular languages that correspond to a function mapping a word to an integer value or infinity. Recently there has been renewed interest in this area, both from a theoretical perspective [14] with max-plus automata, and from a practical point of view with the synthesis of cost register automata [2] for data streams [3]. Within constraint programming, automata constraints were introduced in [18] and in [8,15], the latter also computing an integer value from a word.

This paper focusses on the $g\_\text{SURF}\_\sigma(X, R)$ families of time-series constraints with g being either `Max` or `Sum`, and with σ being one of the 22 patterns of [5], as they were reported to be the most difficult in the recent work of [4]. Each constraint of one of the two families restricts R to be the result of applying the aggregator g to the sum of the elements corresponding to the occurrences of a pattern σ [3] in an integer sequence X, which is called a *time series* and corresponds to measurements taken over time. These constraints play an important role in modelling power systems [10]. If the measured values correspond to the power input/output, then the surface feature `surf` describes the energy

* E. Arafailova is supported by the EU H2020 programme under grant 640954 for the GRACeFUL project. N. Beldiceanu is partially supported by GRACeFUL and by the Gaspard-Monge programme. H. Simonis is supported by Science Foundation Ireland (SFI) under grant numbers SFI/12/RC/2289 and SFI/10/IN.1/I3032.

used/generated during the period of pattern occurence. The `Sum` aggregator imposes a bound on the total energy during all pattern occurences in the time series, the `Max` aggregator is used to limit the maximal energy during a single pattern occurence. Generating time series verifying a set of specific time-series constraints is also useful in different contexts like trace generation, i.e. generating typical energy consumption profiles of a data centre [16,17], or a staff scheduling application, i.e. generating manpower profiles over time subject to work regulations [1,6].

Many constraints of these families are not tractable, thus in order to improve the efficiency of the solving we need to address the combinatorial aspect of time-series constraints. We improve the reasoning for such time-series constraints by identifying implied AMONG constraints. Learning parameters of global constraints like AMONG [9] is a well known method for strengthening constraint models [12,11,19] with the drawback that it is instance specific, so this alternative was not explored here. Taking exact domains into account would lead to filtering algorithms rather than to implied constraints which assume the same minimum/maximum.

While coming up with implied constraints is usually problem specific, the theoretical contribution of this paper is a *unique per family* AMONG implied constraint, *that is valid for all regular expressions* of the time-series constraint catalogue [5] and that covers all the 22 time-series constraints of the corresponding family. Hence, it covers 44 time-series constraints in total. The main focus of this paper is on reusable necessary conditions that can be associated to a class of time-series constraints described with regular expressions. There have been several papers describing progress in propagation of a set of automata and time-series constraints. The techniques described in this paper are only one element required to make such models scale to industrial size.

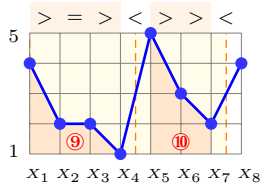
Sec. 2 recalls the necessary background on time-series constraints used in this paper. After introducing several regular expression characteristics, Sec. 3 presents the main contribution, Theorems 1 and 2, while Tables 2 and 3 provide the corresponding derived concrete implied constraints for some subset of the MAX_SURF_σ and the SUM_SURF_σ time-series constraints, respectively, of the time-series constraint catalogue. Finally Sec. 4 systematically evaluates the impact of the derived implied constraints.

2 Time-Series Constraints Background

A time series constraint [7] imposed on a sequence of integer variables $X = \langle X_1, X_2, \ldots, X_n \rangle$ and an integer variable R is described by three main components $\langle g, f, \sigma \rangle$. Let $\mathcal{R}_\Sigma$ denote the set of regular expressions on $\Sigma = \{\text{`<'}, \text{`='}, \text{`>'}\}$. Then, σ is a regular expression in $\mathcal{R}_\Sigma$, that is characterised by two integer constants a_σ and b_σ, whose role is to trim the left and right borders of the regular expression, and $\mathcal{L}_\sigma$ denotes the regular language of σ, while f is a function, called a *feature*. In this paper, we consider only the case when f is `surf`, which will be explained at the end of this paragraph. Finally g is also a function, called

an *aggregator*, that is either `Max` or `Sum`. The *signature* $S = \langle S_1, S_2, \ldots, S_{n-1} \rangle$ of a time series X is defined by the following constraints: $(X_i < X_{i+1} \Leftrightarrow S_i = \text{`<'}) \wedge (X_i = X_{i+1} \Leftrightarrow S_i = \text{`='}) \wedge (X_i > X_{i+1} \Leftrightarrow S_i = \text{`>'})$ for all $i \in [1, n-1]$. If a sub-signature $\langle S_i, S_{i+1}, \ldots, S_j \rangle$ is a maximal word matching σ in the signature of X, then the subseries $\langle X_{i+b_\sigma}, X_{i+b_\sigma+1}, \ldots, X_{j+1-a_\sigma} \rangle$ is called a *σ-pattern* and the subseries $\langle X_i, X_{i+1}, \ldots, X_{j+1} \rangle$ is called an *extended σ-pattern*. The *width* of a σ-pattern is its number of elements. The integer variable R is the aggregation, computed using g, of the list of values of feature f for all σ-patterns in X. The result of applying the `surf` feature to a σ-pattern is the sum of all elements of this σ-pattern. If there is no σ-pattern in X, then R is the *default value*, denoted by $\texttt{def}_{g,f}$, which is $-\infty$, or 0 when g is `Max`, or `Sum`, respectively. A time-series constraint specified by $\langle g, f, \sigma \rangle$ is named as g_f_σ. A time series is *maximal* for $g\_f\_\sigma(X, R)$ if it contains at least one σ-pattern and yields the maximum value of R among all time series of length n that have the same initial domains for the time-series variables.

Example 1. Consider the σ = `DecreasingSequence` = '(>(>|=)*)*>' regular expression and the time series $X = \langle 4, 2, 2, 1, 5, 3, 2, 4 \rangle$ whose signature is '>=><>><'. A σ-pattern, called a *decreasing sequence*, within a time series is a subseries whose signature is a maximal occurrence of σ in the signature of X, and the `surf` feature value of a decreasing sequence is the sum of its elements. The time series X contains two decreasing sequences, namely $\langle 4, 2, 2, 1 \rangle$ and $\langle 5, 3, 2 \rangle$, shown in the figure on the right, of surfaces 9 and 10, respectively. Hence, the aggregation of their surfaces, obtained by using the aggregator `Max`, or `Sum` is 10, or 19 respectively. The corresponding time-series constraints are MAX_SURF_DECREASING_SEQUENCE, and SUM_SURF_DECREASING_SEQUENCE. △



3 Deriving AMONG Implied Constraint

Consider a $g\_f\_\sigma(\langle X_1, X_2, \ldots, X_n \rangle, R)$ time-series constraint with g being either `Sum` or `Max`, with f being the `surf` feature, and with every X_i ranging over the same integer interval domain $[\ell, u]$ such that $u > 0$. For brevity, we do not consider here the case when $u \leq 0$, since it can be handled in a symmetric way. We derive an $\text{AMONG}(\mathcal{N}, \langle X_1, X_2, \ldots, X_n \rangle, \langle \underline{\mathcal{I}}^{\langle \ell,u \rangle}_{\langle g,f,\sigma \rangle}, \underline{\mathcal{I}}^{\langle \ell,u \rangle}_{\langle g,f,\sigma \rangle} + 1, \ldots, \overline{\mathcal{I}}^{\langle \ell,u \rangle}_{\langle g,f,\sigma \rangle} \rangle)$ implied constraint, where:

- For any value of R, $\mathcal{N}$ is an integer variable whose lower bound only depends on R, σ, f, ℓ, u, and n.
- The interval $\mathcal{I}^{\langle \ell,u \rangle}_{\langle g,f,\sigma \rangle} = [\underline{\mathcal{I}}^{\langle \ell,u \rangle}_{\langle g,f,\sigma \rangle}, \overline{\mathcal{I}}^{\langle \ell,u \rangle}_{\langle g,f,\sigma \rangle}]$ is a subinterval of $[\ell, u]$, which is called the *interval of interest of* $\langle g, f, \sigma \rangle$ *wrt* $\langle \ell, u \rangle$ and defined in Sec. 3.1.

Such an AMONG [13] constraint is satisfied if exactly $\mathcal{N}$ variables of $\langle X_1, X_2, \ldots, X_n \rangle$ are assigned a value in $\mathcal{I}^{\langle \ell,u \rangle}_{\langle g,f,\sigma \rangle}$. Before formally describing how to derive this implied constraint, we provide an illustrating example.

Example 2. Consider a MAX_SURF_$\sigma(\langle X_1, X_2, \ldots X_7 \rangle, R)$ time-series constraint with every X_i ranging over the same integer interval domain $[1, 4]$, and with σ being the `DecreasingSequence` regular expression of Ex. 1.

Let us observe what happens when R is fixed, for example, to 18. The table on the right gives the two distinct σ-patterns such that at least one of them appear in every ground time series $X = \langle X_1, X_2, \ldots, X_7 \rangle$ that yields 18 as the value of R. By inspection, we observe that for any ground time series X for which R equals 18, its single σ-pattern contains at least 4 time-series variables whose values are in $[3, 4]$. Hence, we can impose an AMONG$(\mathcal{N}, \langle X_1, X_2, \ldots, X_7 \rangle, \langle 3, 4 \rangle)$ implied constraint with $\mathcal{N} \geq 4$. △

σ-pattern 1	σ-pattern 2
$\langle 4, 3, 3, 3, 3, 2 \rangle$	$\langle 4, 3, 3, 3, 2, 2, 1 \rangle$

We now formalise the ideas presented in Ex. 2 and systematise the way we obtain such an implied constraint even when R is *not initially fixed*.

- Sec. 3.1 introduces five characteristics of a regular expression σ, which will be used to obtain a parameterised implied constraint:
  - the *height of* σ (see Def. 1),
  - the *interval of interest of* $\langle g, f, \sigma \rangle$ *wrt* $\langle \ell, u \rangle$ (see Def. 2),
  - the *maximal value occurrence number of* $v \in \mathbb{Z}$ wrt $\langle \ell, u, n \rangle$ (see Def. 3),
  - the *big width of* σ *wrt* $\langle \ell, u, n \rangle$ (see Def. 4), and
  - the *overlap of* σ *wrt* $\langle \ell, u \rangle$ (see Def. 5).
- Based on these characteristics, Sec. 3.2 presents a systematic way of deriving AMONG implied constraints for the MAX_SURF_σ and the SUM_SURF_σ families of time-series constraints.

3.1 Characteristics of Regular Expressions

To get a unique per family AMONG implied constraint that is valid for any g_SURF_$\sigma(X, R)$ time-series constraint with g being either `Sum` or `Max`, we introduce five characteristics of regular expressions that will be used for parametrising our implied constraint. First, Def. 1 introduces the notion of height of a regular expression, that is needed in Def. 2, which defines the specific range of values on which the implied AMONG constraint focusses on.

Definition 1. *Given a regular expression σ, the* height *of σ, denoted by η_σ, is a function that maps an element of $\mathcal{R}_\Sigma$ to $\mathbb{N}$. It is the smallest difference between the domain upper limit u and the domain lower limit ℓ such that there exists a ground time series over $[\ell, u]$ whose signature has at least one occurrence of σ.*

Example 3. Consider the $\sigma =$ `DecreasingSequence` regular expression of Ex. 1.

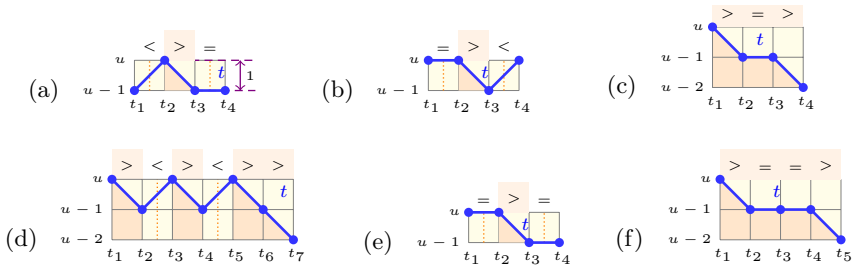


Fig. 1: For all the figures, σ is the `DecreasingSequence` regular expression. A time series t (a) with one σ-pattern such that the difference between its maximum and minimum is 1; (b) with one σ-pattern, which contains a single occurrence of value $u-1$; (c) with one σ-pattern, which contains 2 occurrences of value $u-1$; (d) with the maximum number, 3, of σ-patterns, which all contain one occurrence of value $u-1$, and only one contains an occurrence of value $u-2$; (e) with one σ-pattern, which contains one occurrence of both u and $u-1$; (f) with one σ-pattern, whose width is maximum among all other σ-patterns in ground time series of length 5 over the same domain $[u-2, u]$.

- When $u = \ell$, for any time-series length, there exists a single ground time series t whose signature is a word in the regular language of '=*'. The signature of t contains no occurrences of the '>' symbol, and thus contains no words of $\mathcal{L}_\sigma$ either.
- But when $u - \ell = 1$, there exists, for example, a time series $t = \langle u-1, u, u-1, u-1\rangle$, depicted in Fig. 1a, whose signature '<>=' contains the word '>' of $\mathcal{L}_\sigma$. Hence, the height of σ equals 1. △

Definition 2. *Consider a $g\_f\_\sigma(X, R)$ time-series constraint with X being a time series over an integer interval domain $[\ell, u]$. The* interval of interest *of $\langle g, f, \sigma\rangle$ wrt $\langle \ell, u\rangle$, denoted by $\mathcal{I}^{\langle \ell,u\rangle}_{\langle g,f,\sigma\rangle}$, is a function that maps an element of $\mathcal{T} \times \mathbb{Z} \times \mathbb{Z}$ to $\mathbb{Z} \times \mathbb{Z}$, where $\mathcal{T}$ denotes the set of all time-series constraints, and the result pair of integers is considered as an interval.*

- *The upper limit of $\mathcal{I}^{\langle \ell,u\rangle}_{\langle g,f,\sigma\rangle}$, denoted by $\overline{\mathcal{I}}^{\langle \ell,u\rangle}_{\langle g,f,\sigma\rangle}$, is the largest value in $[\ell, u]$ that can occur in a σ-pattern of a time series over $[\ell, u]$. If such value does not exist, then $\mathcal{I}^{\langle \ell,u\rangle}_{\langle g,f,\sigma\rangle}$ is undefined.*
- *The lower limit of $\mathcal{I}^{\langle \ell,u\rangle}_{\langle g,f,\sigma\rangle}$, denoted by $\underline{\mathcal{I}}^{\langle \ell,u\rangle}_{\langle g,f,\sigma\rangle}$, is the smallest value v in $[\max(\ell, u - \eta_\sigma - 1), u]$ such that for any n in $\mathbb{N}$, the number of occurrences of v in the union of the σ-patterns of any maximal time series for g_f_σ of length n over $[\ell, u]$, is a non-constant function of n. If such v does not exist, then $\underline{\mathcal{I}}^{\langle \ell,u\rangle}_{\langle g,f,\sigma\rangle}$ equals $\overline{\mathcal{I}}^{\langle \ell,u\rangle}_{\langle g,f,\sigma\rangle} - \eta_\sigma$.*

We focus on such intervals of interests because they consist of the largest values appearing in maximal time series for g_f_σ.

Example 4. Consider a $g\_f\_\sigma(X, R)$ time-series constraint with σ being the `DecreasingSequence` regular expression, with f being the `surf` feature, and with X being a time series of length $n \geq 2$ over an integer interval domain $[\ell, u]$ such that $u > 1$ and $u > \ell$. We consider different combinations of triples $\langle g, f, \sigma \rangle$ and their corresponding intervals of interest wrt $\langle \ell, u \rangle$. Note that the value of $\overline{\mathcal{I}}^{\langle \ell,u \rangle}_{\langle g,f,\sigma \rangle}$ depends only on σ, ℓ, and u and not on g and f. The largest value appearing in the σ-patterns of X is u, and thus $\overline{\mathcal{I}}^{\langle \ell,u \rangle}_{\langle g,f,\sigma \rangle} = u$. We compute the value of $\underline{\mathcal{I}}^{\langle \ell,u \rangle}_{\langle g,f,\sigma \rangle}$ wrt two time-series constraints:

- Let g be the `Max` aggregator.
  - If $u - \ell = 1$, then any σ-pattern of X has a signature '>', i.e. contains only two elements. Then, the maximum value of R is reached for a time series t that contains the $\langle u, u-1 \rangle$ σ-pattern. The rest of the variables of t are assigned any value, e.g. all other variables have a value of u. Such a time series t for the length 4 is shown in Fig. 1b. Further, for any v in $[\ell, u]$, the number of occurrences of v in the union of the σ-patterns of t is at most 1, which is a constant, and does not depend on n. By definition $\underline{\mathcal{I}}^{\langle \ell,u \rangle}_{\langle g,f,\sigma \rangle} = \overline{\mathcal{I}}^{\langle \ell,u \rangle}_{\langle g,f,\sigma \rangle} - \eta_\sigma = u - 1$.
  - If $u - \ell > 1$, then any maximal time series t for g_f_σ contains a single σ-pattern whose signature is in the language of '>=*>'. If, for example, $n = 4$, then t has $n - 2 = 2$ time-series variables with the values $u - 1$, which is depicted Fig. 1c. In addition, the σ-pattern of t has a single occurrence of the value $u - 2$. Hence, $\underline{\mathcal{I}}^{\langle \ell,u \rangle}_{\langle g,f,\sigma \rangle} = u - 1$.
- Let g be the `Sum` aggregator.
  Any maximal time series t for g_f_σ contains $\lfloor \frac{n}{2} \rfloor$ σ-patterns, which contains u and $u-1$, and at most one of them has the value $u-2$. Such a time series t for the length $n = 7$ is depicted in Fig. 1d. Hence, $\underline{\mathcal{I}}^{\langle \ell,u \rangle}_{\langle g,f,\sigma \rangle} = u-1$. $\triangle$

The next characteristic, we introduce, is a function of ℓ, u and n related to the maximum number of value occurrences in a σ-pattern.

Definition 3. *Consider a regular expression σ, and a time series X of length n over an integer interval domain $[\ell, u]$. The* maximum value occurrence number *of v in $\mathbb{Z}$ wrt $\langle \ell, u, n \rangle$, denoted by $\mu_\sigma^{\langle \ell,u,n \rangle}(v)$, is a function that maps an element of $\mathcal{R}_\Sigma \times \mathbb{Z} \times \mathbb{Z} \times \mathbb{N}^+ \times \mathbb{Z}$ to $\mathbb{N}$. It equals the maximum number of occurrences of the value v in one σ-pattern of X.*

Example 5. Consider the σ = `DecreasingSequence` regular expression and a time series X of length $n \geq 2$ over an integer interval domain $[\ell, u]$ such that $u > \ell$. We compute the maximum value occurrence number of v in $\mathbb{Z}$ wrt $\langle \ell, u, n \rangle$. If v is not in $[\ell, u]$, then $\mu_\sigma^{\langle \ell,u,n \rangle}(v) = 0$. Hence, we focus on the case when $v \in [\ell, u]$.

- If $u - \ell = 1$, then any σ-pattern of X has a signature '>', and thus it may have at most one occurrence of any value v in $[\ell, u]$. Hence, for any v in $[\ell, u]$, $\mu_\sigma^{\langle \ell,u,n \rangle}(v) = 1$.

- If $u - \ell > 1$, then we consider two subsets of $[\ell, u]$:
  - For either v in the set $\{\ell, u\}$, the value of $\mu_\sigma^{\langle \ell,u,n\rangle}(v)$ is 1, since in any σ-pattern the lower and upper limits of the domain, namely ℓ and u, can appear at most once, as it illustrated in Fig. 1e for the length $n = 4$.
  - For any v in $[\ell + 1, u - 1]$, the value of $\mu_\sigma^{\langle \ell,u,n\rangle}(v)$ is $\max(1, n - 2)$, since v can occur at most $n - 2$ times in a σ-pattern of X. The time series in Fig. 1c has a single σ-pattern, namely $\langle t_1, t_2, t_3, t_4\rangle$, which has $n - 2 = 4 - 2 = 2$ occurrences of the value $u - 1$. △

The next characteristic, we introduce, is the largest width of a σ-pattern in a time series.

Definition 4. *Consider a regular expression σ, and a time series X of length n over an integer interval domain $[\ell, u]$. The* big width *of σ wrt $\langle \ell, u, n\rangle$, denoted by $\beta_\sigma^{\langle \ell,u,n\rangle}$, is a function that maps an element of $\mathcal{R}_\Sigma \times \mathbb{Z} \times \mathbb{Z} \times \mathbb{N}^+$ to $\mathbb{N}$. It equals the maximum width of a σ-pattern in X. If X cannot have any σ-patterns, then $\beta_\sigma^{\langle \ell,u,n\rangle}$ is 0.*

Example 6. Consider the σ = `DecreasingSequence` regular expression and a time series X of length n over an integer interval domain $[\ell, u]$.

- If $n \leq 1$, then X cannot have any σ-patterns, since a minimum width σ-pattern contains at least two elements. Hence, $\beta_\sigma^{\langle \ell,u,n\rangle} = 0$.
- If $u - \ell = 0$, then, as it was shown in Ex. 3, no word of $\mathcal{L}_\sigma$ can appear in the signature of any ground time series over $[\ell, u]$, and thus X cannot have any σ-patterns. Hence, $\beta_\sigma^{\langle \ell,u,n\rangle} = 0$.
- If $u - \ell = 1$ and $n \geq 2$, then any σ-pattern of X has a signature '>'. The width of such a σ-pattern is 2. Hence, $\beta_\sigma^{\langle \ell,u,n\rangle} = 2$.
- If $u - \ell > 1$ and $n \geq 2$, then there exists a word in $\mathcal{L}_\sigma$ that is also in the language of '>=*>' and whose length is $n - 1$. This word is the signature of some ground time series t of length n over $[\ell, u]$, which contains a single σ-pattern of width n. Such a time series t for the length $n = 5$ is illustrated in Fig. 1f. The width of a σ-pattern cannot be greater than n, thus $\beta_\sigma^{\langle \ell,u,n\rangle} = n$. △

The last characteristic is the notion of maximum overlap of a regular expression wrt an integer interval domain. It will be used for deriving an implied AMONG constraint when the aggregator of a considered time-series constraint is `Sum`.

Definition 5. *Consider a regular expression σ and an integer interval domain $[\ell, u]$. The* overlap of σ wrt $[\ell, u]$, *denoted by $o_\sigma^{\langle \ell,u\rangle}$, is the maximum number of time-series variables that belong simultaneously to two extended σ-patterns of a time series among all time series over $[\ell, u]$. If such maximum number does not exist, then $o_\sigma^{\langle \ell,u\rangle}$ is undefined.*

Example 7. Consider the σ = `DecreasingSequence` regular expression and an interval $[\ell, u]$ with $u > \ell$. For any time series over $[\ell, u]$, any of its two extended σ-patterns have no time-series variables in common, thus $o_\sigma^{\langle \ell,u\rangle} = 0$. △

σ	η_σ	$\mu_\sigma^{\langle \ell,u,n\rangle}(v)$	$\beta_\sigma^{\langle \ell,u,n\rangle}$	$o_\sigma^{\langle \ell,u\rangle}$
' >><<>> '	2	$\begin{cases} 1, \text{ if } v \in \{\ell, \ell+1, u-1, u\} \\ 2, \text{ if } v \in [\ell+2, u-2] \end{cases}$	3	3
'>'	1	$1, \forall v \in [\ell, u]$	2	1
'(>(>|=)*)*>'	1	$\begin{cases} 1, \text{ if } v \in \{u, \ell\} \\ \max(1, n-2), \text{ if } v \in [\ell+1, u-1] \end{cases}$	$\begin{cases} 2, \text{ if } u-\ell=1 \\ n, \text{ Otherwise} \end{cases}$	0
'(>(>|=)*)*><((<|=)*<)*'	1	$\begin{cases} 0, \text{ if } v = u \\ n-3, \text{ if } v \in [\ell+1, u-1] \\ 1, \text{ if } v = \ell \end{cases}$	$\begin{cases} 1, \text{ if } u-\ell=1 \\ n-2, \text{ Otherwise} \end{cases}$	1
'<(<|=)* (>|=)*>'	1	$\begin{cases} 0, \text{ if } v = \ell \\ n-2, \text{ if } v \in [\ell+1, u] \end{cases}$	$n-2$	1
'(<>)$^+$(< | <>)|(><)$^+$(> | ><)'	1	$\lfloor \frac{n-1}{2} \rfloor$, $\forall v \in [\ell, u]$	$n-2$	$\begin{cases} 0, \text{ if } u-\ell=1 \\ 1, \text{ Otherwise} \end{cases}$

Table 1: For every regular expression σ, $[\ell, u]$ is an integer interval domain, and n is a time series length, such that there is at least one ground time series of length n over $[\ell, u]$ whose signature contains at least one occurrence of σ. Then, η_σ is the height of σ, $\mu_\sigma^{\langle \ell,u,n\rangle}(v)$ is the maximum value occurrence number of $v \in [\ell, u]$ wrt $\langle \ell, u, n\rangle$, $\beta_\sigma^{\langle \ell,u,n\rangle}$ is the the big width of σ wrt $\langle \ell, u, n\rangle$, and $o_\sigma^{\langle \ell,u\rangle}$ is the overlap of σ wrt $\langle \ell, u\rangle$.

Table 1 gives the values of the four characteristics of regular expressions for some regular expressions of [5], while Tables 2 and 3 provide the intervals of interest for 12 time-series constraints.

3.2 Deriving an AMONG Implied Constraint for the MAX_SURF_σ and the SUM_SURF_σ Families

Consider a $g\_f\_\sigma(\langle X_1, X_2, \ldots, X_n\rangle, R)$ time-series constraint with every X_i ranging over the same integer interval domain $[\ell, u]$, with f being the `surf` feature, and with g being either `Max` or `Sum`. Our goal is to estimate a lower bound on $\mathcal{N}$, which is the number of time-series variables in the σ-patterns of $\langle X_1, X_2, \ldots, X_n\rangle$ that must be assigned a value in the interval of interest $\mathcal{I}_{\langle g,f,\sigma\rangle}^{\langle \ell,u\rangle}$ of $\langle g, f, \sigma\rangle$ wrt $\langle \ell, u\rangle$, in order to satisfy the $g\_f\_\sigma(\langle X_1, X_2, \ldots, X_n\rangle, R)$ constraint. Theorems 1 and 2 present such inequality for the cases when g is `Max`, and `Sum`, respectively, using the four characteristics introduced in Sec. 3.1. Ex. 8 first conveys the intuition behind Thm. 1.

Example 8. Consider a $g\_f\_\sigma(X, R)$ time-series constraint with g being `Max`, with f being `surf`, with σ being the `DecreasingSequence` regular expression, and with X being a time series of length $n = 9$ over the integer interval domain $[\ell, u] = [0, 4]$. Let us assign R to the value 24, and let us compute a lower bound on $\mathcal{N}$, the number of variables of X that must be assigned a value from $\mathcal{I}_{\langle g,f,\sigma\rangle}^{\langle \ell,u\rangle}$, which is $[3, 4]$ as it was shown in Ex. 4. Our aim is to show that for a σ-pattern in X, its number of time-series variables in $[3, 4]$ can be estimated as the difference between the value of the surface of this σ-pattern and some other value that is a function of σ, ℓ, u and n. In order to obtain this value,

we construct a time series t of length $\beta_\sigma^{\langle \ell,u,n\rangle} = 9$ satisfying all the following conditions:

1. The number of time-series variables of t that are assigned to the value $\overline{\mathcal{I}}_{\langle g,f,\sigma\rangle}^{\langle \ell,u\rangle}$ equals $\mu_\sigma^{\langle \ell,u,n\rangle}(\overline{\mathcal{I}}_{\langle g,f,\sigma\rangle}^{\langle \ell,u\rangle}) = \mu_\sigma^{\langle 0,4,9\rangle}(4) = 1$.
2. The number of time-series variables of t that are assigned to the value $\underline{\mathcal{I}}_{\langle g,f,\sigma\rangle}^{\langle \ell,u\rangle}$, which is $\overline{\mathcal{I}}_{\langle g,f,\sigma\rangle}^{\langle \ell,u\rangle} - 1$, equals $\mu_\sigma^{\langle \ell,u,n\rangle}(\underline{\mathcal{I}}_{\langle g,f,\sigma\rangle}^{\langle \ell,u\rangle}) = \mu_\sigma^{\langle 0,4,9\rangle}(3) = n - 2 = 7$.
3. The rest of the time-series variables of t, namely $n - \mu_\sigma^{\langle \ell,u,n\rangle}(\overline{\mathcal{I}}_{\langle g,f,\sigma\rangle}^{\langle \ell,u\rangle}) - \mu_\sigma^{\langle \ell,u,n\rangle}(\underline{\mathcal{I}}_{\langle g,f,\sigma\rangle}^{\langle \ell,u\rangle}) = 1$ time-series variable, is assigned to the value $\underline{\mathcal{I}}_{\langle g,f,\sigma\rangle}^{\langle \ell,u\rangle} - 1 = 2$.

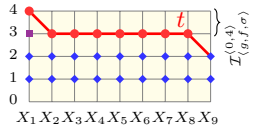
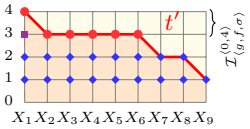


Figure on the left illustrates a ground time series t of length 9 over $[0, 4]$ satisfying all the three conditions. By construction, the sum of elements of t is greater than or equal to the surface of any σ-pattern of X. Furthermore, for any σ-pattern of X, its number of time-series variables whose values are in $[3, 4]$ is not greater than the number of such time-series variables of t.

Figure above on the left contains three type of points: circled, squared and diamond-shaped points; thus our goal is to evaluate the number of circles. The value of X_i is one plus the number of squared and diamond-shaped points under the point corresponding to X_i. Hence, the sum of all elements of t can be viewed as the total number of circled, squared and diamond-shaped points. Furthermore, the number of circles is the difference between the total number of points and the number of squared points, namely 27 minus 19, which is 8.

For any σ-pattern of X, its corresponding number of squared and diamond-shaped points is at most 19. Then, its number of time-series variables whose values are in $[3, 4]$ can be estimated as the surface of the σ-pattern minus 19. Hence, when the surface of the σ-pattern is 24, a lower bound on $\mathcal{N}$ is 5. Figure on the right gives an example of a ground time series t' of length 9 over $[0, 4]$ that contains a σ-pattern with a surface of 24. This σ-pattern has $6 \geq 5$ values in $[3, 4]$, which agrees with our computed lower bound. △



Theorem 1. *Consider a* $g\_f\_\sigma(X, R)$ *time-series constraint with* $g = \texttt{Max}$, $f = \texttt{surf}$ *and* X *being a time series of length* n *over an integer interval domain* $[\ell, u]$; *then* AMONG$(\mathcal{N}, X, \mathcal{I})$ *is an implied constraint, where* $\mathcal{N}$ *is restricted by*

$$\mathcal{N} \geq R - \max(0, \underline{\mathcal{I}} - 1) \cdot \beta - \sum_{v \in [\underline{\mathcal{I}}+1, \overline{\mathcal{I}}]} \mu_\sigma^{\langle \ell,u,n\rangle}(v) \cdot (v - \underline{\mathcal{I}}), \qquad (1)$$

where β *(resp.* $\mathcal{I}$*) is shorthand for* $\beta_\sigma^{\langle \ell,u,n\rangle}$ *(resp.* $\mathcal{I}_{\langle g,f,\sigma\rangle}^{\langle \ell,u\rangle}$*), and* $\underline{\mathcal{I}}$ *(resp.* $\overline{\mathcal{I}}$*) denotes the lower (resp. upper) limit of interval* $\mathcal{I}$.

Proof We show that the right-hand side of the stated inequality is a lower bound on the number of time-series variables of a σ-pattern whose values are in $\mathcal{I}$, and the surface of the σ-pattern is R. In order to prove the lower bound on $\mathcal{N}$, we first compute a lower bound on the number $\mathcal{N}^{\underline{\mathcal{I}}}$ of time-series variables of the σ-pattern whose values are $\underline{\mathcal{I}}$, which is the smallest value of interval $\mathcal{I}$. We assume that for every $v > \underline{\mathcal{I}}$ in $\mathcal{I}$, the number of occurrences of v in the σ-pattern equals some $\mathcal{N}^v$. Note that the number of time-series variables in any σ-pattern is not greater than $\beta = \beta_\sigma^{\langle \ell,u,n \rangle}$. We state the following inequality:

$$R \leq \underbrace{\mathcal{N}^{\underline{\mathcal{I}}} \cdot \max(0, \underline{\mathcal{I}})}_{A} + \underbrace{\sum_{v \in [\underline{\mathcal{I}}+1, \overline{\mathcal{I}}]} \mathcal{N}^v \cdot \max(0, v)}_{B} + \underbrace{\max(0, \underline{\mathcal{I}} - 1) \cdot (\beta - \mathcal{N}^{\underline{\mathcal{I}}} - \sum_{v \in [\underline{\mathcal{I}}+1, \overline{\mathcal{I}}]} \mathcal{N}^v)}_{C}, \tag{2}$$

where A, B, and C correspond to the sums of elements of the σ-pattern that equal $\underline{\mathcal{I}}$, are in $\mathcal{I}$ and are greater than $\underline{\mathcal{I}}$, and are outside $\mathcal{I}_{\langle g,f,\sigma \rangle}^{\langle \ell,u \rangle}$ respectively. From Inequality (2) we obtain the following lower bound on $\mathcal{N}^{\underline{\mathcal{I}}}$:

$$\mathcal{N}^{\underline{\mathcal{I}}} \geq R - \sum_{v \in [\underline{\mathcal{I}}+1, \overline{\mathcal{I}}]} \mathcal{N}^v \cdot \max(0, v) - \max(0, \underline{\mathcal{I}} - 1) \cdot (\beta - \sum_{v \in [\underline{\mathcal{I}}+1, \overline{\mathcal{I}}]} \mathcal{N}^v). \tag{3}$$

In order to obtain a lower bound on $\mathcal{N}$ from the known lower bound on $\mathcal{N}^{\underline{\mathcal{I}}}$, we add $\sum_{v \in [\underline{\mathcal{I}}+1, \overline{\mathcal{I}}]} \mathcal{N}^v$ to both sides of Inequality (3). Further, we regroup some terms in Inequality (3), we eliminate $\sum_{v \in [\underline{\mathcal{I}}+1, \overline{\mathcal{I}}]} \mathcal{N}^v$ in the right-hand side of Inequality (3) by replacing it with $\sum_{v \in [\underline{\mathcal{I}}+1, \overline{\mathcal{I}}]} \mu_\sigma^{\langle \ell,u,n \rangle}(v)$, and obtain the inequality of the theorem. □

Example 9. Consider the $g\_f\_\sigma(\langle X_1, X_2, \ldots, X_n \rangle, R)$ time-series constraint, with g being `Sum`, with f being `surf`, and with every X_i (with $i \in [1, n]$) ranging over the same domain $[\ell, u]$ with $u > 1$ and $u - \ell > 1$. We illustrate the derivation of AMONG implied constraints for two regular expressions.

- Consider the σ = `DecreasingSequence` regular expression and $n \geq 2$. In Ex. 4, we computed the interval of interest of MAX_SURF_σ wrt $\langle \ell, u \rangle$, which is $[u-1, u]$. In Ex. 5, we showed that $\mu_\sigma^{\langle \ell,u,n \rangle}(\ell) = \mu_\sigma^{\langle \ell,u,n \rangle}(u) = 1$, and for every value v in $[\ell+1, u-1]$, we have that $\mu_\sigma^{\langle \ell,u,n \rangle}(v)$ equals $\max(1, n-2)$. Finally, in Ex. 6 we demonstrated that $\beta_\sigma^{\langle \ell,u,n \rangle} = n$. By Thm. 1, we can impose the AMONG$(\mathcal{N}, X, \langle u-1, u \rangle)$ implied constraint with $\mathcal{N} \geq R - \mu_\sigma^{\langle \ell,u,n \rangle}(u) - \max(0, \underline{\mathcal{I}}_{\langle g,f,\sigma \rangle}^{\langle \ell,u \rangle} - 1) \cdot \beta_\sigma^{\langle \ell,u,n \rangle} = R - 1 - \max(0, u-2) \cdot n$. Turning back to Ex. 8 we observe that, in the obtained implied constraint, the term '1'

σ	$\mathcal{I}^{\langle \ell,u \rangle}_{\langle \text{MAX,SURF},\sigma \rangle}$	LB
‘ >><<>> ’	$[u-2,u]$	$R - \max(0,u-3)\cdot 3 - 3$
‘>’	$[u-1,u]$	$R - \max(0,u-2)\cdot 2 - 1$
‘(>(>|=)*)*>’	$[u-1,u]$	$R - \max(0,u-2)\cdot n - 1$
‘(>(>|=)*)*><((<|=)*<)*’	$[u-1,u-1]$	$R - \max(0,u-2)\cdot (n-2)$
‘<(<|=)* (>|=)*>’	$[u,u]$	$R - \max(0,u-1)\cdot (n-2)$
‘(<>)$^+$(< | <>)|(><)$^+$(> | ><)’	$[u-1,u]$	$R - \max(0,u-2)\cdot (n-2) - \lfloor \frac{n-1}{2} \rfloor$

Table 2: Regular expression σ, the corresponding interval of interest of MAX_SURF_$\sigma(X,R)$ wrt an integer interval domain $[\ell,u]$ such that $u > 1$ and $u - \ell > 1$, and the lower bound LB on the parameter of the derived AMONG implied constraint. The value LB is obtained from a generic formula, which is parameterised by characteristics of regular expressions. The sequence X is supposed to be long enough to contain at least one σ-pattern.

corresponds to the number of squared points, and the term ‘$\max(0,u-2)\cdot n$’ to the number of diamond-shaped points. The derived lower bound on $\mathcal{N}$ also appears in the third row of Table 2.

- Consider the $\sigma = \texttt{Peak} =$ ‘<(<|=)* (>|=)*>’ regular expression whose values of a_σ and b_σ both equal 1, and $n \geq 3$. The maximum value in $[\ell,u]$ that appears in a σ-pattern is u. In addition, any maximal time series for $\langle g,f,\sigma \rangle$ contains a single σ-pattern whose values are all the same and equal u. Hence, the interval of interest of $\langle g,f,\sigma \rangle$ wrt $\langle \ell,u \rangle$ is $[u,u]$. Since both a_σ and b_σ equal 1, the smallest value in $[\ell,u]$ may not be in any σ-pattern and $\mu_\sigma^{\langle \ell,u,n \rangle}(\ell) = 0$. For any value $v \in [\ell+1,u]$, we have $\mu_\sigma^{\langle \ell,u,n \rangle}(v) = n-2$. By Thm. 2, we impose an AMONG$(\mathcal{N},\langle X_1,X_2,\ldots,X_n \rangle, \langle u \rangle)$ implied constraint with $\mathcal{N} \geq R - \max(0,u-1)\cdot(n-2)$. The derived lower bound on $\mathcal{N}$ also appears in the fifth row of Table 2. △

Table 2 gives for 6 regular expressions of [5] the corresponding intervals of interest of MAX_SURF_σ constraints wrt some integer interval domain $[\ell,u]$ such that $u > 1 \ \wedge \ u - \ell > 1$, as well as the lower bound LB on the parameter $\mathcal{N}$ of the derived AMONG constraint for time series that may have at least one σ-pattern.

Theorem 2. *Consider a* $g\_f\_\sigma(X,R)$ *time-series constraint with* $g = \texttt{Sum}$, $f = \texttt{surf}$ *and* X *being a time series of length* n *over an integer interval domain* $[\ell,u]$*; then* AMONG$(\mathcal{N},X,\mathcal{I})$ *is an implied constraint, where* $\mathcal{N}$ *is restricted by*

$$\begin{aligned} \mathcal{N} \geq \ & R - \max\left(0,\underline{\mathcal{I}}-1\right)\cdot\Big(n - a_\sigma - b_\sigma + (p_o-1)\cdot\max(0,o_\sigma^{\langle \ell,u \rangle} - a_\sigma - b_\sigma)\Big) \\ & - \sum_{v \in [\underline{\mathcal{I}}+1,\overline{\mathcal{I}}]} \mu_\sigma^{\langle \ell,u,n \rangle}(v)\cdot p_o \cdot (v - \underline{\mathcal{I}}) \\ & - \ (p_o-1)\cdot\max(0,o_\sigma^{\langle \ell,u \rangle} - a_\sigma - b_\sigma)\ , \end{aligned} \tag{4}$$

where $\mathcal{I}$ is shorthand for $\mathcal{I}^{\langle \ell,u \rangle}_{\langle g,f,\sigma \rangle}$, $\underline{\mathcal{I}}$ (resp. $\overline{\mathcal{I}}$) denotes the lower (resp. upper) limit of $\mathcal{I}$, and p_o is 1 if every maximal time series has a single σ-pattern, and is the maximal number of σ-patterns in a time series of length n, otherwise.

Proof. To prove Thm. 2 we consider a time series with $p \geq 1$ σ-patterns, where σ-pattern i (with $i \in [1,p]$) has a width of ω_i and a surface of R_i, and where $R = \sum_{i \in [1,p]} R_i$. The proof consists of two steps:

1. First, for each σ-pattern i (with $i \in [1,p]$), we compute the minimum number $\mathcal{N}_i$ of time-series variables that must be assigned to a value within the interval of interest $\mathcal{I}$, in order to reach a surface of R_i.
2. Second, we take the sum of $\mathcal{N}_i$, and minimise the obtained value, which, in the end, will be a minimum value for $\mathcal{N}$.

First Step. We use Inequality (1) of Thm. 1 for a subseries X' of X of length $\omega'_i = \omega_i + a_\sigma + b_\sigma$, knowing that X' has a single σ-pattern and $\beta_\sigma^{\langle \ell,u,n \rangle}$ is ω_i. Then, by Thm. 1, we obtain the following estimation of $\mathcal{N}_i$:

$$\mathcal{N}_i \geq R_i - \omega_i \cdot \max(0, \underline{\mathcal{I}} - 1) - \sum_{v \in [\underline{\mathcal{I}}+1, \overline{\mathcal{I}}]} (v - \underline{\mathcal{I}}) \cdot \mu_\sigma^{\langle \ell,u,\omega'_i \rangle}(v). \tag{5}$$

Second Step. We obtain the minimum value of $\mathcal{N}$, by taking the sum of the derived minimum values for $\mathcal{N}_i$ over all the values of i:

$$\mathcal{N} = \sum_{i=1}^{p} \mathcal{N}_i \geq \sum_{i=1}^{p} (R_i - A_i - B_i) - C = R - \sum_{i=1}^{p} A_i - \sum_{i=1}^{p} B_i - C, \tag{6}$$

where for any $i \in [1,p]$, $A_i = \omega_i \cdot \max(0, \underline{\mathcal{I}} - 1)$ and $B_i = \sum_{v \in [\underline{\mathcal{I}}+1, \overline{\mathcal{I}}]} \mu_\sigma^{\langle \ell,u,\omega'_i \rangle}(v) \cdot (v - \underline{\mathcal{I}})$, and $C = (p-1) \cdot \max(0, o_\sigma^{\langle \ell,u \rangle} - a_\sigma - b_\sigma)$. The terms A_i and B_i come from Inequality (5) and the term C is used because some variables may belong to two σ-patterns: in order to not count them twice we subtract a correction term. Let A (resp. B) denote $\sum_{i=1}^{p} A_i$ (resp. $\sum_{i=1}^{p} B_i$). In order to satisfy Condition 6, we need to find the upper bounds on the sum $A+B+C$ by choosing the value of p, and the sum of σ-patterns lengths. We consider two cases, but any additional information may be used for a more accurate estimation of these parameters:

- [EVERY MAXIMAL TIME SERIES HAS A SINGLE σ-PATTERN] Then, the maximum value of $A+B+C$ is reached for p being 1, and $\sum_{i=1}^{p} \omega_i$ being $n - b_\sigma - a_\sigma$. It implies that for any $v \in [\underline{\mathcal{I}}^{\langle \ell,u \rangle}_{\langle g,f,\sigma \rangle} + 1, \overline{\mathcal{I}}^{\langle \ell,u \rangle}_{\langle g,f,\sigma \rangle}]$, the value of $\sum_{i \in [1,p]} \mu_\sigma^{\langle \ell,u,\omega'_i \rangle}(v)$ equals $\mu_\sigma^{\langle \ell,u,n \rangle}(v)$.
- [THERE IS AT LEAST ONE MAXIMAL TIME SERIES WITH MORE THAN ONE σ-PATTERN] We give an overestimation: we assign the value of p to its maximum value, which depends on σ, the value of $\sum_{i=1}^{p} \omega_i$ is overestimated by $n -$

σ	$\mathcal{I}^{\langle \ell,u \rangle}_{\langle \text{SUM},\text{SURF},\sigma \rangle}$	LB
‘ >><<>> ’	$[u-2,u]$	$R - \max(0,u-3)\cdot(n-3) - 3\cdot\lfloor\frac{n-3}{3}\rfloor$
‘>’	$[u-1,u]$	$R - \max(0,u-2)\cdot(2\cdot n-2) - (2\cdot n-3)$
‘(>(>|=)*)*>’	$[u-1,u]$	$R - \max(0,u-2)\cdot n - \lfloor\frac{n}{2}\rfloor$
‘(>(>|=)*)*><((<|=)*<)*’	$[u-1,u-1]$	$R - \max(0,u-2)\cdot(n-2)$
‘<(<|=)* (>|=)*>’	$[u,u]$	$R - \max(0,u-1)\cdot(n-2)$
‘(<>)$^+$(< | <>)|(><)$^+$(> | ><)’	$[u-1,u]$	$R - \max(0,u-2)\cdot(n-2) - \lfloor\frac{n-1}{2}\rfloor$

Table 3: Regular expression σ, the corresponding interval of interest of SUM_SURF_$\sigma(X,R)$ wrt an integer interval domain $[\ell,u]$ such that $u>1$ and $u-\ell>1$, and the lower bound LB on the parameter of the derived AMONG implied constraint. The value LB is obtained from a generic formula, which is parameterised by characteristics of regular expressions. The sequence X is supposed to be long enough to contain at least one σ-pattern.

$a_\sigma - b_\sigma + (p_o-1)\cdot\max(0, o_\sigma^{\langle \ell,u \rangle} - a_\sigma - b_\sigma)$, and the value of $\sum_{i\in[1,p]} \mu_\sigma^{\langle \ell,u,\omega'_i \rangle}(v)$ is overestimated by $\mu_\sigma^{\langle \ell,u,n \rangle}(v)\cdot p_o$.

Hence, we obtain a lower bound for $\mathcal{N}$, which is the right hand side of the inequality stated by Thm. 2. □

Example 10. Consider the $g\_f\_\sigma(\langle X_1, X_2, \ldots, X_n\rangle, R)$ time-series constraint, with g being `Sum`, with f being `surf` and with every X_i (with $i\in[1,n]$) ranging over the same domain $[\ell,u]$ with $u>1$ and $u-\ell>1$. We illustrate the derivation of AMONG implied constraints for two regular expressions.

- Consider the σ = `DecreasingSequence` regular expression and $n\geq 2$. In Ex. 4, we found that the interval of interest of $\langle g,f,\sigma\rangle$ wrt $\langle \ell,u\rangle$ is $[u-1,u]$, and in Ex. 5, we showed that $\mu_\sigma^{\langle \ell,u,n \rangle}(\ell) = \mu_\sigma^{\langle \ell,u,n \rangle}(u) = 1$, and for every value v in $[\ell+1,u-1]$, we have that $\mu_\sigma^{\langle \ell,u,n \rangle}(v)$ equals $\max(1,n-2)$. Every maximal time series for SUM_SURF_σ contains the maximum number of σ-patterns. Hence, in this case, the value of p_o equals the maximum number of decreasing sequences in a time series of length n, which is $\lfloor\frac{n}{2}\rfloor$. By Thm. 2, we impose an AMONG$(\mathcal{N},\langle X_1,X_2,\ldots,X_n\rangle,\langle u-1,u\rangle)$ implied constraint with $\mathcal{N}\geq R - \lfloor\frac{n}{2}\rfloor - \max(0,u-2)\cdot n$. The derived lower bound on $\mathcal{N}$ also appears in the third row of Table 3.
- Consider the σ = `Peak` = ‘<(<|=)* (>|=)*>’ regular expression and $n\geq 3$. The maximum value in $[\ell,u]$ that occurs in a σ-pattern is u. In addition, any maximal time series for $\langle g,f,\sigma\rangle$ contains a single σ-pattern whose values are all the same and equal u. Hence, the interval of interest of $\langle g,f,\sigma\rangle$ wrt $\langle \ell,u\rangle$ is $[u,u]$, and the value of p_o equals 1. We showed in Ex. 9 that $\mu_\sigma^{\langle \ell,u,n \rangle}(\ell) = 0$ and for any $v\in[\ell+1,u]$, we have $\mu_\sigma^{\langle \ell,u,n \rangle}(v) = n-2$. The value of $o_\sigma^{\langle \ell,u \rangle}$ equals 1. By Thm. 2, we impose an AMONG$(\mathcal{N},\langle X_1,\ X_2,\ldots,X_n\rangle,\ \langle u\rangle)$ implied constraint with $\mathcal{N}\geq R - \max(0,u-1)\cdot(n-2)$. The derived lower bound on $\mathcal{N}$ also appears in the fifth row of Table 3. △

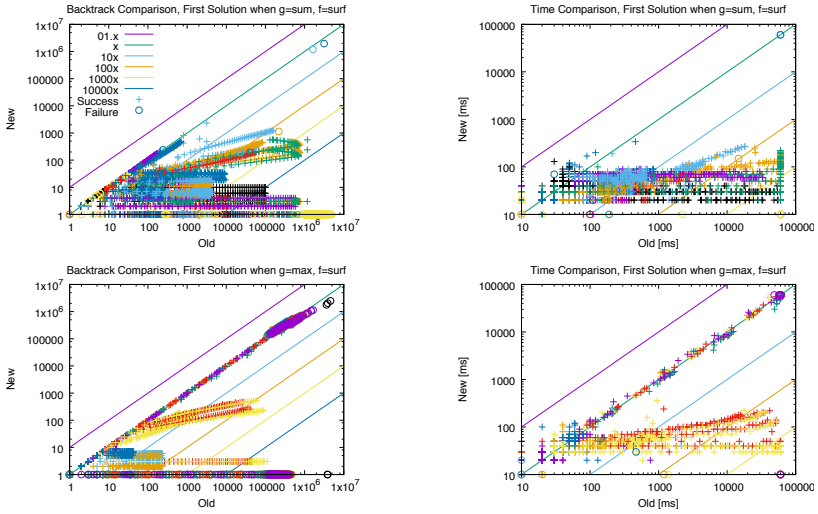


Fig. 2: Comparing backtrack count and runtime of the g_f_σ time-series constraints for previous best results (old) and new method for finding the first solution or proving infeasibility for time series of length 50 and domain $[0, 5]$. Colours of markers indicate the regular expression, the cross (resp. circle) marker type indicates success (resp. failure/timeout).

Table 3 gives for 6 regular expressions of [5] the corresponding intervals of interest of SUM_SURF_σ constraints wrt some integer interval domain $[\ell, u]$ such that $u > 1 \wedge u - \ell > 1$, as well as the lower bound LB on the parameter $\mathcal{N}$ of the derived AMONG constraint for time series that may have at least one σ-pattern.

4 Evaluation

The intended use case is a problem where we learn parameters for a conjunction of many time-series constraints from data, and use this conjunction to create new time-series that are "similar" to the existing ones. An example would be electricity production data for a day [10], in half hour periods (48 values), or manpower levels per week over a year (52 values). To solve the conjunction, we need strong propagation for each individual constraint. We therefore evaluate the impact of the implied constraint on both execution time and the number of backtracks for the time-series constraints of the MAX_SURF_σ and the SUM_SURF_σ families for which a glue constraint [4] exists, which are 38 out of 44 time-series constraints of the two families. These families of constraints were the most difficult to solve in the experiments reported in [4].

In the experiments for both families, we consider a single g_SURF_$\sigma(X, R)$ time-series constraint with g being either `Sum` or `Max`, for which we first systematically try out all potential values of the parameter R, and then either

find a solution by assigning the X_i or prove infeasibility. We compare the best (Combined) approach from the recent work [4] to the new method, adding the implied AMONG constraint on every suffix of $X = \langle X_1, X_2, \ldots, X_n \rangle$, and also a *preprocessing procedure*. The preprocessing procedure is a useful, if minor, contribution of the paper for 8 out of 38 of the constraints in the families studied. The purpose of this procedure is to find all feasible values of R, when σ is such that any σ-pattern has all values being the same. Such values of R must satisfy the following constraint:

$$R = \texttt{def}_{g,f} \quad \vee \quad \left(\exists V \in [\ell', u'] \ \beta_\sigma^{\langle \ell, u, n \rangle} \cdot V \geq R \quad \wedge \quad R \bmod V = 0 \right),$$

where ℓ' and u' are the smallest and the largest value, respectively, that can occur in a σ-pattern over $[\ell, u]$.

Since the implied constraints are precomputed offline, posting one implied constraint takes a *constant time*, and the time and space complexity of the preprocessing procedure does not exceed the size of the domain of R, which is $O(n \cdot (u - \ell))$.

Fig. 2 presents the results for the SUM_SURF_σ (upper plots) and the MAX_SURF_σ (lower plots) time-series constraints, where X is a time series of length 50 over the domain $[0, 5]$, when the goal is to find, for each value of R, the first solution or prove infeasibility. This corresponds to our main use case, where we want to construct time series with fixed R values. Our static search routine enumerates the time-series variables X_i from left to right, starting with the smallest value in the domain. Results for the backtrack count are on the left, results for the execution time on the right. We use log scales on both axes, replacing a zero value by one in order to allow plotting. A timeout of 60 seconds was imposed. We see that the implied constraints reduce backtracks by up to a factor exceeding 10,000 and runtime by up to a factor of 1,000, and they divide the total execution time of terminated instances by a factor of 5 and 45 times when g is `Max` and `Sum`, respectively. All experiments were run on a 2014 iMac 4 GHz *i*7 using SICStus Prolog.

The results for the case $g =$ `Sum` are better than for the case $g =$ `Max` because the aggregator `Sum` allows summing the surfaces of several σ-patterns, whereas for the `Max` aggregator, R is the surface of a single σ-pattern, the surfaces of other σ-patterns, if any, are absorbed.

5 Conclusion

In summary, based on 4 regular expression characteristics, we have defined a *single per family generic implied constraint* for all constraints of the MAX_SURF_σ and SUM_SURF_σ families. The experimental results showed a good speed up in the number of backtracks and the time spent for the SUM_SURF_σ family.

References

1. O. Zeynep Akşin, Mor Armony, and Vijay Mehrotra. The modern call center: A multi-disciplinary perspective on operations management research. *Production and Operations Management*, 16(6):665–688, 2007.
2. Rajeev Alur, Loris D'Antoni, Jyotirmoy V. Deshmukh, Mukund Raghothaman, and Yifei Yuan. Regular functions and cost register automata. In *28th Annual ACM/IEEE Symposium on Logic in Computer Science, LICS 2013, New Orleans, LA, USA, June 25-28, 2013*, pages 13–22. IEEE Computer Society, 2013.
3. Rajeev Alur, Dana Fisman, and Mukund Raghothaman. Regular programming for quantitative properties of data streams. In Peter Thiemann, editor, *Programming Languages and Systems - 25th European Symposium on Programming, ESOP 2016*, volume 9632 of *Lecture Notes in Computer Science*, pages 15–40. Springer, 2016.
4. Ekaterina Arafailova, Nicolas Beldiceanu, Mats Carlsson, Pierre Flener, María Andreína Francisco Rodríguez, Justin Pearson, and Helmut Simonis. Systematic derivation of bounds and glue constraints for time-series constraints. In Michel Rueher, editor, *CP 2016*, volume 9892 of *LNCS*, pages 13–29. Springer, 2016.
5. Ekaterina Arafailova, Nicolas Beldiceanu, Rémi Douence, Mats Carlsson, Pierre Flener, María Andreína Francisco Rodríguez, Justin Pearson, and Helmut Simonis. Global constraint catalog, volume ii, time-series constraints. *CoRR*, abs/1609.08925, 2016.
6. Ekaterina Arafailova, Nicolas Beldiceanu, Rémi Douence, Pierre Flener, María Andreína Francisco Rodríguez, Justin Pearson, and Helmut Simonis. Time-series constraints: Improvements and application in CP and MIP contexts. In Claude-Guy Quimper, editor, *CP-AI-OR 2016*, volume 1713 of *LNCS*, pages 18–34. Springer, 2016.
7. Nicolas Beldiceanu, Mats Carlsson, Rémi Douence, and Helmut Simonis. Using finite transducers for describing and synthesising structural time-series constraints. *Constraints*, 21(1):22–40, January 2016. Journal fast track of CP 2015: summary on p. 723 of LNCS 9255, Springer, 2015.
8. Nicolas Beldiceanu, Mats Carlsson, and Thierry Petit. Deriving filtering algorithms from constraint checkers. In Mark Wallace, editor, *CP 2004*, volume 3258 of *LNCS*, pages 107–122. Springer, 2004.
9. Nicolas Beldiceanu and Evelyne Contejean. Introducing global constraints in CHIP. *Mathl. Comput. Modelling*, 20(12):97–123, 1994.
10. Nicolas Beldiceanu, Georgiana Ifrim, Arnaud Lenoir, and Helmut Simonis. Describing and generating solutions for the EDF unit commitment problem with the ModelSeeker. In Christian Schulte, editor, *CP 2013*, volume 8124 of *LNCS*, pages 733–748. Springer, 2013.
11. Christian Bessière, Remi Coletta, Emmanuel Hébrard, George Katsirelos, Nadjib Lazaar, Nina Narodytska, Claude-Guy Quimper, and Toby Walsh. Constraint Acquisition via Partial Queries. In *IJCAI 2013*, page 7, Beijing, China, June 2013.
12. Christian Bessière, Remi Coletta, and Thierry Petit. Learning Implied Global Constraints. In *IJCAI 2007*, pages 50–55, Hyderabad, India, 2007.
13. Christian Bessière, Emmanuel Hebrard, Brahim Hnich, Zeynep Kiziltan, and Toby Walsh. Among, common and disjoint constraints. In *Recent Advances in Constraints, Joint ERCIM/CoLogNET International Workshop on Constraint Solving and Constraint Logic Programming, CSCLP 2005*, pages 29–43, 2005.
14. Thomas Colcombet and Laure Daviaud. Approximate comparison of functions computed by distance automata. *Theory Comput. Syst.*, 58(4):579–613, 2016.

15. Sophie Demassey, Gilles Pesant, and Louis-Martin Rousseau. A `Cost-Regular` based hybrid column generation approach. *Constraints*, 11(4):315–333, 2006.
16. Lieven Eeckhout, Koen De Bosschere, and Henk Neefs. Performance analysis through synthetic trace generation. In *2000 ACM/IEEE Intl. Symp. Performance Analysis Syst. Software*, pages 1–6, 2000.
17. Lars Kegel, Martin Hahmann, and Wolfgang Lehner. Template-based time series generation with loom. In *EDBT/ICDT Workshops 2016, Bordeaux, France*, 2016.
18. Gilles Pesant. A regular language membership constraint for finite sequences of variables. In Mark Wallace, editor, *CP 2004*, volume 3258 of *LNCS*, pages 482–495. Springer, 2004.
19. Émilie Picard-Cantin, Mathieu Bouchard, Claude-Guy Quimper, and Jason Sweeney. Learning parameters for the sequence constraint from solutions. In Michel Rueher, editor, *CP 2016*, volume 9892 of *LNCS*, pages 405–420. Springer, 2016.
20. Marcel Paul Schützenberger. On the definition of a family of automata. *Information and Control*, 4:245–270, 1961.