## Systematic Derivation of Bounds and Glue Constraints for Time-Series Constraints

Ekaterina Arafailova, Nicolas Beldiceanu, Mats Carlsson, Pierre Flener, M. Andreína Francisco R., Justin Pearson, and Helmut Simonis

7th September 2016
CP 2016

## Contents



## Example of a Time-Series Constraint

Constrain the maximum of the minima of the valleys in the time series $\langle 5,5,6,4,6,6,4,2,4,4,1,1,1,1,3,0\rangle$.

## Example of a Time-Series Constraint

Constrain the maximum of the minima of the valleys in the time series $\langle 5,5,6,4,6,6,4,2,4,4,1,1,1,1,3,0\rangle$.


## Example of a Time-Series Constraint

Constrain the maximum of the minima of the valleys in the time series $\langle 5,5,6,4,6,6,4,2,4,4,1,1,1,1,3,0\rangle$.


## Example of a Time-Series Constraint

Constrain the maximum of the minima of the valleys in the time series $\langle 5,5,6,4,6,6,4,2,4,4,1,1,1,1,3,0\rangle$.


## Example of a Time-Series Constraint

Constrain the maximum of the minima of the valleys in the time series $\langle 5,5,6,4,6,6,4,2,4,4,1,1,1,1,3,0\rangle$.


A subsequence $\left\langle X_{i}, \ldots, X_{j}\right\rangle$ of $\left\langle X_{0}, \ldots, X_{m}\right\rangle$ is a valley if the signature of $\left\langle X_{i-1}, \ldots, X_{j+1}\right\rangle$ is a maximal word matching ' $>(>\mid=)^{*}(<\mid=)^{*}<^{\prime}$.

## Example of a Time-Series Constraint

Constrain the maximum of the minima of the valleys in the time series $\langle 5,5,6,4,6,6,4,2,4,4,1,1,1,1,3,0\rangle$.


A subsequence $\left\langle X_{i}, \ldots, X_{j}\right\rangle$ of $\left\langle X_{0}, \ldots, X_{m}\right\rangle$ is a valley if the signature of $\left\langle X_{i-1}, \ldots, X_{j+1}\right\rangle$ is a maximal word matching ' $>(>\mid=)^{*}(<\mid=)^{*}<$ '.

## Compositional Time-Series Definition by Multiple Layers of Functions


max_min_valley $(\langle 5,5,6,4,6,6,4,2,4,4,1,1,1,1,3,0\rangle, 4)$ holds

## Space of Time-Series Constraints



## Space of Time-Series Constraints



## Main Question

How to go from a compositional constraint definition to compositional combinatorial objects that can be used in different contexts, such as CP, MIP, local search, and data mining?

## Main Question

How to go from a compositional constraint definition to compositional combinatorial objects that can be used in different contexts, such as CP, MIP, local search, and data mining?

Time-series constraints are very numerous: we cannot afford to consider each of them separately.

## Applications of Time-Series Constraints

- Analysis of output of electric power stations
[CP 2013] over multiple days.
- Power management for
[Computing, 2016] large-scale-distributed systems.
- Trace analysis for Internet Service Provider.
- Anomaly detection and error correction in building data.


## Synthesising Time-Series Constraints

Use a transducer:
[Constraints, 2016]

- Input sequence: Handle the matching aspect
- Output sequence: Handle the computational aspect


## Synthesising Time-Series Constraints

Use a transducer:
[Constraints, 2016]

- Input sequence: Handle the matching aspect
- Output sequence: Handle the computational aspect


Transducer for valley

## Synthesising Time-Series Constraints

Use a transducer:
[Constraints, 2016]

- Input sequence: Handle the matching aspect
- Output sequence: Handle the computational aspect


Transducer for valley

## Synthesising Time-Series Constraints

Use a transducer:
[Constraints, 2016]

- Input sequence: Handle the matching aspect
- Output sequence: Handle the computational aspect


Transducer for valley


Automaton for max_min_valley

## Synthesising Time-Series Constraints

Use a transducer:
[Constraints, 2016]

- Input sequence: Handle the matching aspect
- Output sequence: Handle the computational aspect


Transducer for valley


Automaton for max_min_valley

$$
\frac{5}{s} \underset{\text { out }}{=} \frac{5}{s} \underset{\text { out }}{<} \frac{6}{s} \xrightarrow[\text { out }]{>} \frac{4}{r} \xrightarrow[\text { found }]{<} \frac{6}{t} \xrightarrow[\text { maybe }_{a}]{=} \frac{6}{t} \xrightarrow[\text { out }_{a}]{>} \frac{4}{r} \xrightarrow[\text { maybe }_{b}]{>} \frac{2}{r} \underset{\text { found }}{<} \frac{4}{t}
$$

## Implementing Time-Series Constraints

- CP: Reformulation of automata with accumulators*
- MIP: Linear reformulation of automata with accumulators**
* [Constraints, 2005], ** [CPAIOR 2016]

Goal: Address the combinatorial aspect of time-series constraints.
Method: Exploit the compositional nature of time-series constraints at the combinatorial level.

## Contents



## Compositional Sharp Bounds: Contents



## Context and Goal

Consider a g_f_ $\sigma\left(\left\langle X_{0}, \ldots, X_{m}\right\rangle, N\right)$ time-series constraint, for a regular expression $\sigma$, a feature f , and an aggregator g , with every $X_{i}$ ranging over the same integer interval $[a, b]$.

## Context and Goal

Consider a g_f_ $\sigma\left(\left\langle X_{0}, \ldots, X_{m}\right\rangle, N\right)$ time-series constraint, for a regular expression $\sigma$, a feature f , and an aggregator g , with every $X_{i}$ ranging over the same integer interval $[a, b]$.

Goal: Derive formulae $\boldsymbol{\beta}_{\mathrm{f}, \mathrm{g}}^{\mathbf{u}}$ and $\boldsymbol{\beta}_{\mathrm{f}, \mathrm{g}}^{\boldsymbol{\ell}}$ that are parametrised by $\sigma$ and yield respectively sharp upper and lower bounds on the value of $N$.

## Compositional Sharp Bounds: Contents



## Computing a Sharp Upper Bound for the pair $\langle$ min, max

Consider a max_min_ $\sigma\left(\left\langle X_{0}, \ldots, X_{m}\right\rangle, N\right)$ time-series constraint with every $X_{i}$ ranging over [a, $b$ ].

Goal: Compute a sharp upper bound on the value of $N$.

## Computing a Sharp Upper Bound for the pair $\langle$ min, max〉

Consider a max_min_ $\sigma\left(\left\langle X_{0}, \ldots, X_{m}\right\rangle, N\right)$ time-series constraint with every $X_{i}$ ranging over $[a, b]$.

Goal: Compute a sharp upper bound on the value of $N$.

## Method:

1. Introduce the characteristics of a regular expression $\sigma$ : the minimal difference between the domain upper bound $b$ and the minima of the subseries, which is called the shift of $\sigma$ and denoted by $\Delta_{\sigma}$.

## Computing a Sharp Upper Bound for the pair $\langle\mathrm{min}$, max $\rangle$

Consider a max_min_ $\sigma\left(\left\langle X_{0}, \ldots, X_{m}\right\rangle, N\right)$ time-series constraint with every $X_{i}$ ranging over $[a, b]$.

Goal: Compute a sharp upper bound on the value of $N$.

## Method:

1. Introduce the characteristics of a regular expression $\sigma$ : the minimal difference between the domain upper bound $b$ and the minima of the subseries, which is called the shift of $\sigma$ and denoted by $\Delta_{\sigma}$.
2. Explore the connection between this characteristics and the sharp upper bound on the value of $N$.

## Compositional Sharp Bounds: Contents



## Computing a Sharp Upper Bound for max_min_valley

Consider the max_min_valley $\left(\left\langle X_{0}, \ldots, X_{8}\right\rangle, N\right)$ time-series constraints with every $X_{i}$ ranging over $[a, b]=[3,6]$.

## Shift of valley and Its Connection to a Sharp Upper Bound

- The shift of valley is the smallest possible difference between $b$ and the minima of valleys, namely 1 and is reached, for example, for valleys whose signature is ' $><$ '.


## Shift of valley and Its Connection to a Sharp Upper Bound

- The shift of valley is the smallest possible difference between $b$ and the minima of valleys, namely 1 and is reached, for example, for valleys whose signature is ' $><$ '.
- The minimum of any valley is at least $\Delta_{\text {valley }}$, which is 1 , away from $b$, which is 6 , thus $N \leq b-\Delta_{\text {valley }}=6-1=5$.


## Shift of valley and Its Connection to a Sharp Upper Bound

- The shift of valley is the smallest possible difference between $b$ and the minima of valleys, namely 1 and is reached, for example, for valleys whose signature is ' $><$ '.
- The minimum of any valley is at least $\Delta_{\text {valley }}$, which is 1 , away from $b$, which is 6 , thus $N \leq b-\Delta_{\text {valley }}=6-1=5$.
- This bound is reached for some ground time series.


## Shift of valley and Its Connection to a Sharp Upper Bound

- The shift of valley is the smallest possible difference between $b$ and the minima of valleys, namely 1 and is reached, for example, for valleys whose signature is ' $><$ '.
- The minimum of any valley is at least $\Delta_{\text {valley }}$, which is 1 , away from $b$, which is 6 , thus $N \leq b-\Delta_{\text {valley }}=6-1=5$.
- This bound is reached for some ground time series.

$$
\begin{aligned}
& N=b-\Delta_{\text {valley }}=6-1=5 \\
& \begin{array}{l}
b=6 \\
4------------\infty
\end{array} \\
& 3^{x_{0}} \quad x_{1} \quad x_{2} \quad x_{3} \quad x_{4} \quad x_{5} \quad x_{6} \quad x_{7} \quad x_{8} \\
& S:=<><=\gg<
\end{aligned}
$$

## Compositional Sharp Bounds: Contents



## Sharp Upper Bound for the General Case

Theorem

Consider a max_min_ $\sigma\left(\left\langle X_{0}, \ldots, X_{m}\right\rangle, N\right)$ time-series constraint, where $\sigma$ is one of the 22 regular expressions and every $X_{i}$ ranges over $[a, b]$, then the sharp upper bound on the value of $N$ is

$$
\boldsymbol{\beta}_{\min , \max }^{\mathrm{u}}=b-\Delta_{\sigma}
$$

## Compositional Bounds (Summary)



## Contents



## Linking the Prefixes and Suffixes of a Sequence

For the ngroup $\left(\left\langle X_{0}, \ldots, X_{m}\right\rangle, \mathcal{S}, N\right)$ constraint, $N$ is the number of maximal subsequences of $\left\langle X_{0}, \ldots, X_{m}\right\rangle$ whose values are all in $\mathcal{S}$.

## Linking the Prefixes and Suffixes of a Sequence

For the ngroup $\left(\left\langle X_{0}, \ldots, X_{m}\right\rangle, \mathcal{S}, N\right)$ constraint, $N$ is the number of maximal subsequences of $\left\langle X_{0}, \ldots, X_{m}\right\rangle$ whose values are all in $\mathcal{S}$.

The ngroup $(\langle 1,0,1,1,0,0\rangle,\{1\}, 2)$ constraint holds.

## Linking the Prefixes and Suffixes of a Sequence

For the ngroup $\left(\left\langle X_{0}, \ldots, X_{m}\right\rangle, \mathcal{S}, N\right)$ constraint, $N$ is the number of maximal subsequences of $\left\langle X_{0}, \ldots, X_{m}\right\rangle$ whose values are all in $\mathcal{S}$.

The ngroup $(\langle 1,0,1,1,0,0\rangle,\{1\}, 2)$ constraint holds.


## Linking the Prefixes and Suffixes of a Sequence

For the ngroup $\left(\left\langle X_{0}, \ldots, X_{m}\right\rangle, \mathcal{S}, N\right)$ constraint, $N$ is the number of maximal subsequences of $\left\langle X_{0}, \ldots, X_{m}\right\rangle$ whose values are all in $\mathcal{S}$.

The ngroup $(\langle 1,0,1,1,0,0\rangle,\{1\}, 2)$ constraint holds.

not just a sum $(2 \neq 2+1) \ldots$

## Linking the Prefixes and Suffixes of a Sequence

For the ngroup $\left(\left\langle X_{0}, \ldots, X_{m}\right\rangle, \mathcal{S}, N\right)$ constraint, $N$ is the number of maximal subsequences of $\left\langle X_{0}, \ldots, X_{m}\right\rangle$ whose values are all in $\mathcal{S}$.

The ngroup $(\langle 1,0,1,1,0,0\rangle,\{1\}, 2)$ constraint holds.

| $\operatorname{ngroup}(\langle 1,0,1,1,0,0\rangle,\{1\}, 2)$ |  |  |  |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: |
| 1, | 0, | 1, | 1, | 0, | 0 |

## not just a sum $(2 \neq 2+1)$...how to compute the correction term -1 ?

## Glue Matrix for the ngroup Constraint



Automaton for ngroup

## Glue Matrix for the ngroup Constraint



Automaton for ngroup

## Glue Matrix for the ngroup Constraint



Automaton for ngroup


Glue Matrix for ngroup

## Glue Matrix for the ngroup Constraint



Automaton for ngroup


Glue Matrix for ngroup

$$
\operatorname{ngroup}(\langle 1,0,1,1,0,0\rangle,\{1\}, 2)
$$

$\xrightarrow[\text { ngroup }(\langle 1,0,1\rangle,\{1\}, 2)]{$| 1, | 0, | 1, | $1, \quad 0,$ | 0 |
| :---: | :---: | :---: | :---: | :---: | :---: |
| $t, \quad u, \quad$ | $t, \quad \mathrm{u},$ | $\mathrm{u}, \quad t, \quad t,$ | $t$ |  |
|  ngroup $(\langle 0,0,1\rangle,\{1\}, 1)$ |  |  |  |  |$}$

## Compositional Glue Matrices for Time-Series Constraints

Goal: For a transducer, derive a glue matrix that is parametrised by a feature and an aggregator.

## Generic Glue Matrix for steady _sequence $=$ ' $={ }^{+}$'



Generic Automaton for g_f_steady_sequence

## Generic Glue Matrix for steady _sequence $=$ ' $={ }^{+}$



Generic Automaton for g_f_steady_sequence

Generic Glue Matrix for g_f_steady_sequence

## Corresponding Compositional Glue Matrices

$$
\left\{\begin{array}{l}
C \leftarrow \text { default } \\
D \leftarrow 0 \\
R \leftarrow R+C
\end{array}\right\}
$$

## Contents



## Synergy of Sharp Bounds and Glue Matrices

Consider a g_f_ $\sigma\left(\left\langle X_{0}, \ldots, X_{m}\right\rangle, N\right)$ time-series constraint where all $X_{i}$ range over the same domain.

- Add sharp bounds for every prefix and every reversed suffix of $X_{0}, \ldots, X_{m}$.
- Impose glue-matrix implied constraints on every prefix and every reversed suffix of $X_{0}, \ldots, X_{m}$.

The combination of sharp bounds and glue matrices improves propagation.

## Conclusion: clear separation between three levels

- Language: compositional description of constraints with aggregator g , feature f , regular expression $\sigma$
- Abstract Combinatorial Objects (transducers and formulae)
- Transducers for each regular expression $\sigma$
- Parametrised bounds for each pair g,f
- Parametrised glue matrices for each regular expression $\sigma$ (22)
- Multiple usages (synthesised code: ATEX (154100), Prolog (135267))
- CP: enhanced propagation (bounds, glue matrices)
- Local search: constant time probing (glue matrices)
- Data mining: range of variation of feature

Separate definition from usage, be compositional at all levels

Forthcoming in late 2016:
Automatically synthesised catalogue of time-series constraints, with all transducers, constraints, bounds, glue matrices, linear invariants, and MiniZinc reformulations.

## Forthcoming in late 2016:

Automatically synthesised catalogue of time-series constraints, with all transducers, constraints, bounds, glue matrices, linear invariants, and MiniZinc reformulations.

Thank you for your attention!


围 Arafailova，Ekaterina et al．（2016）．＂Time－series constraints： Improvements and application in CP and MIP contexts＂．In： CPAIOR 2016．Ed．by Claude－Guy Quimper．Vol．9676．LNCS． Springer，pp．18－34．
Beldiceanu，Nicolas et al．（2005）．＂Reformulation of global constraints based on constraints checkers＂．In：Constraints 10．4， pp．339－362．
围 Beldiceanu，Nicolas et al．（2013）．＂Describing and generating solutions for the EDF unit commitment problem with the ModelSeeker＂．In：CP 2013．Ed．by Christian Schulte．Vol． 8124. LNCS．Springer，pp．733－748．
围 Beldiceanu，Nicolas et al．（2016a）．＂Towards energy－proportional Clouds partially powered by renewable energy＂．In：Computing， p．20．url：https：／／hal．inria．fr／hal－01340318．
盏 Beldiceanu，Nicolas et al．（2016b）．＂Using finite transducers for describing and synthesising structural time－series constraints＂． In：Constraints 21．1，pp．22－40．

