(In-)Variant Synthesis for Probabilistic Programs

[Immature Idea]

Mingshuai Chen

—Joint work with Shenghua Feng, Joost-Pieter Katoen, and 😍—

Largest **I**nternational **C**olloquium at **S**öllerhaus (LICS 2020)

Lehrstuhl für Informatik 2 Softwaremodellierung und Verifikation | RWTH AACHEN UNIVERSITY

Mathematical Ingredients : SDP & SOS

Semidefinite Programming

Definition (SDP [Wolkowicz et al., 2012])

Semidefinite programming (SDP) refers to optimization problems of the form

$$\begin{array}{lll} \underset{X}{\texttt{minimize}} & C \bullet X & \\ \texttt{subject to} & A_i \bullet X = \mathbf{b}_i & \text{for } i = 1, \ldots, m \\ & X \succeq 0 & \end{array}$$

with the variable $X \in \mathcal{S}^n$, and (given) problem parameters $\mathbf{b} \in \mathbb{R}^m$ and $C, A_i \in \mathcal{S}^n$.

Semidefinite Programming

Definition (SDP [Wolkowicz et al., 2012])

Semidefinite programming (SDP) refers to optimization problems of the form

$$\begin{array}{lll} \underset{X}{\texttt{minimize}} & C \bullet X & \\ \texttt{subject to} & A_i \bullet X = \mathbf{b}_i & \text{for } i = 1, \ldots, m \\ & X \succeq 0 & \end{array}$$

with the variable $X \in \mathcal{S}^n$, and (given) problem parameters $\mathbf{b} \in \mathbb{R}^m$ and $C, A_i \in \mathcal{S}^n$.

- A generalization of LP : $\mathbf{x} \geq 0 \longrightarrow X \succeq 0$.

Semidefinite Programming

Definition (SDP [Wolkowicz et al., 2012])

Semidefinite programming (SDP) refers to optimization problems of the form

$$\begin{array}{lll} \underset{X}{\texttt{minimize}} & C \bullet X & \\ \texttt{subject to} & A_i \bullet X = \mathbf{b}_i & \text{for } i = 1, \ldots, m \\ & X \succeq 0 & \end{array}$$

with the variable $X \in \mathcal{S}^n$, and (given) problem parameters $\mathbf{b} \in \mathbb{R}^m$ and $C, A_i \in \mathcal{S}^n$.

- A generalization of LP : $\mathbf{x} \geq 0 \longrightarrow X \succeq 0$.
- The generalization preserves *convexity* $\Longrightarrow$ SDPs admit polynomial-time algorithms, e.g., the interior-point methods, albeit numerical.

Sum of Squares

Definition (SOS Polynomial)

A polynomial $p(\mathbf{x}) \in \mathbb{R}[\mathbf{x}]$ is a *sum of squares* (SOS), if there are polynomials $q_1(\mathbf{x}), \ldots, q_m(\mathbf{x})$ s.t.

$$p(\mathbf{x}) = \sum_{i=1}^{m} q_i^2(\mathbf{x}).$$

Sum of Squares

Definition (SOS Polynomial)

A polynomial $p(\mathbf{x}) \in \mathbb{R}[\mathbf{x}]$ is a *sum of squares* (SOS), if there are polynomials $q_1(\mathbf{x}), \ldots, q_m(\mathbf{x})$ s.t.

$$p(\mathbf{x}) = \sum_{i=1}^{m} q_i^2(\mathbf{x}).$$

- $p(\mathbf{x})$ is an SOS $\implies$ $p(\mathbf{x}) \geq 0$.

Sum of Squares

Definition (SOS Polynomial)

A polynomial $p(\mathbf{x}) \in \mathbb{R}[\mathbf{x}]$ is a *sum of squares* (SOS), if there are polynomials $q_1(\mathbf{x}), \ldots, q_m(\mathbf{x})$ s.t.

$$p(\mathbf{x}) = \sum_{i=1}^{m} q_i^2(\mathbf{x}).$$

- $p(\mathbf{x})$ is an SOS $\implies$ $p(\mathbf{x}) \geq 0$.
- $p(\mathbf{x})$ is an SOS $\iff$ there exist $\mathbf{z} \mathrel{\widehat{=}} [1, \mathbf{x}_0, \mathbf{x}_1, \mathbf{x}_0\mathbf{x}_1, \ldots, \mathbf{x}_n^d]$ and $Q \succeq 0$ s.t.

$$p(\mathbf{x}) = \mathbf{z}^\mathsf{T} Q \mathbf{z}.$$

Sum of Squares

Definition (SOS Polynomial)

A polynomial $p(\mathbf{x}) \in \mathbb{R}[\mathbf{x}]$ is a *sum of squares* (SOS), if there are polynomials $q_1(\mathbf{x}), \ldots, q_m(\mathbf{x})$ s.t.

$$p(\mathbf{x}) = \sum_{i=1}^{m} q_i^2(\mathbf{x}).$$

- $p(\mathbf{x})$ is an SOS $\implies$ $p(\mathbf{x}) \geq 0$.
- $p(\mathbf{x})$ is an SOS $\iff$ there exist $\mathbf{z} \mathrel{\widehat{=}} [1, \mathbf{x}_0, \mathbf{x}_1, \mathbf{x}_0\mathbf{x}_1, \ldots, \mathbf{x}_n^d]$ and $Q \succeq 0$ s.t.

$$p(\mathbf{x}) = \mathbf{z}^\mathsf{T} Q \mathbf{z}.$$

⇒ Finding an SOS can be encoded as solving an SDP problem.

Sum of Squares

Example ([Sankaranarayanan et al., 2016])

$p(x,y) \mathrel{\widehat{=}} 2x^4 + 2x^3y - x^2y^2 + 5y^4$ is an SOS, since for instance

$$p(x,y) = \begin{bmatrix} x^2 \\ y^2 \\ xy \end{bmatrix}^{\mathsf{T}} \begin{bmatrix} 2 & -3 & 1 \\ -3 & 5 & 0 \\ 1 & 0 & 5 \end{bmatrix} \begin{bmatrix} x^2 \\ y^2 \\ xy \end{bmatrix} \quad \text{and}$$

$$\begin{bmatrix} 2 & -3 & 1 \\ -3 & 5 & 0 \\ 1 & 0 & 5 \end{bmatrix} = L^{\mathsf{T}}L \quad \text{with} \quad L = \frac{1}{\sqrt{2}} \begin{bmatrix} 2 & -3 & 1 \\ 0 & 1 & 3 \end{bmatrix} \quad \text{hence}$$

$$p(x,y) = \frac{1}{2}\left(2x^2 - 3y^2 + xy\right)^2 + \frac{1}{2}\left(y^2 + 3xy\right)^2 .$$

SOS Relaxation

Lemma (SOS Relaxation [Feng et al., 2017])

The following statements hold (with $\triangleright_i \in \{\geq, =\}$):

1. $f(\mathbf{x}) \geq 0 \implies g(\mathbf{x}) \geq 0$ *holds if* $g(\mathbf{x}) - u \cdot f(\mathbf{x})$ *is an SOS for some* $u \in \mathbb{R}_{\geq 0}$.
2. $f(\mathbf{x}) = 0 \implies g(\mathbf{x}) \geq 0$ *holds if* $g(\mathbf{x}) - v \cdot f(\mathbf{x})$ *is an SOS for some* $v \in \mathbb{R}$.
3. $\bigwedge_i f_i(\mathbf{x}) \triangleright_i 0 \implies g(\mathbf{x}) \geq 0$ *holds if* $g(\mathbf{x}) - \sum_i r_i \cdot f_i(\mathbf{x})$ *is an SOS for some* $r_i \in \mathbb{R}$*;*
   If $\triangleright_i$ *is* $\geq$*, it is additionally required that* $r_i \geq 0$.

SOS Relaxation

Lemma (SOS Relaxation [Feng et al., 2017])

The following statements hold (with $\triangleright_i \in \{\geq, =\}$):

1. $f(\mathbf{x}) \geq 0 \implies g(\mathbf{x}) \geq 0$ *holds if* $g(\mathbf{x}) - u \cdot f(\mathbf{x})$ *is an SOS for some* $u \in \mathbb{R}_{\geq 0}$.
2. $f(\mathbf{x}) = 0 \implies g(\mathbf{x}) \geq 0$ *holds if* $g(\mathbf{x}) - v \cdot f(\mathbf{x})$ *is an SOS for some* $v \in \mathbb{R}$.
3. $\bigwedge_i f_i(\mathbf{x}) \triangleright_i 0 \implies g(\mathbf{x}) \geq 0$ *holds if* $g(\mathbf{x}) - \sum_i r_i \cdot f_i(\mathbf{x})$ *is an SOS for some* $r_i \in \mathbb{R}$*;*
   If $\triangleright_i$ *is* $\geq$*, it is additionally required that* $r_i \geq 0$.

- A negation on the LHS can be eliminated using De Morgan's laws.

SOS Relaxation

Lemma (SOS Relaxation [Feng et al., 2017])

The following statements hold (with $\rhd_i \in \{\geq, =\}$):

1. $f(\mathbf{x}) \geq 0 \implies g(\mathbf{x}) \geq 0$ *holds if* $g(\mathbf{x}) - u \cdot f(\mathbf{x})$ *is an SOS for some* $u \in \mathbb{R}_{\geq 0}$.
2. $f(\mathbf{x}) = 0 \implies g(\mathbf{x}) \geq 0$ *holds if* $g(\mathbf{x}) - v \cdot f(\mathbf{x})$ *is an SOS for some* $v \in \mathbb{R}$.
3. $\bigwedge_i f_i(\mathbf{x}) \rhd_i 0 \implies g(\mathbf{x}) \geq 0$ *holds if* $g(\mathbf{x}) - \sum_i r_i \cdot f_i(\mathbf{x})$ *is an SOS for some* $r_i \in \mathbb{R}$*; If* $\rhd_i$ *is* $\geq$*, it is additionally required that* $r_i \geq 0$.

- A negation on the LHS can be eliminated using De Morgan's laws.
- A disjunction, e.g., $\phi \vee \psi \implies \chi$ can be equivalently split as

$$(\phi \implies \chi) \wedge (\psi \implies \chi).$$

Supermartingales Witnessing Almost-Sure Termination

Variant Rule for Loops

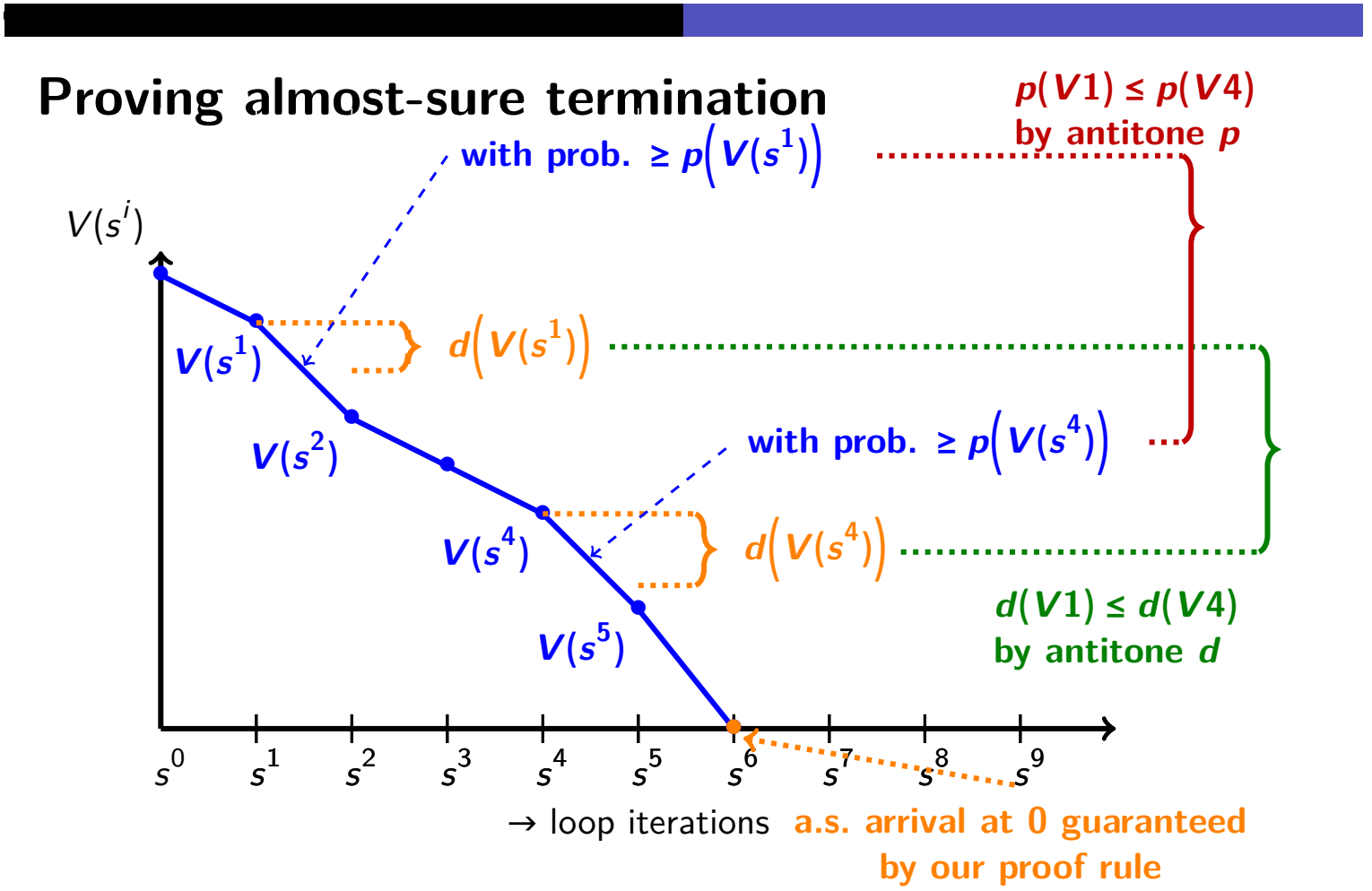
Theorem (Variant Rule for Loops [McIver et al., 2018])

Let $I, G \subseteq \Sigma$ be predicates; let $V\colon \Sigma \to \mathbb{R}_{\geq 0}$ be a non-negative real-valued function not necessarily bounded; let $p\colon \mathbb{R}_{\geq 0} \to (0,1]$ (for "probability") and $d\colon \mathbb{R}_{\geq 0} \to \mathbb{R}_{>0}$ (for "decrease") be fixed functions with antitonicity on strictly positive arguments; and let Com be a pGCL *program. Suppose the following four conditions hold :*

1. *I is a standard invariant of* $\texttt{while}(G)\{Com\}$, *and*
2. $G \wedge I \implies V > 0$, *and*
3. $\forall R \in \mathbb{R}_{>0}\colon p(R) \cdot [G \wedge I \wedge V = R] \leq \mathrm{wp}.Com.[V \leq R - d(R)]$, *and*
4. $\forall H \in \mathbb{R}_{>0}\colon [G \wedge I] \cdot (H \ominus V) \leq \mathrm{wp}.Com.(H \ominus V)$, *with* $H \ominus V \mathrel{\widehat{=}} \max\{H - V, 0\}$.

Then, $[I] \leq \mathrm{wp}.\texttt{while}(G)\{Com\}.1$.

Variant Rule for Loops



Variant Rule for Loops

Example (The Escaping Spline [McIver et al., 2018])

$$\texttt{while}(x > 0)\{q := 1/x+1; \{x := 0\}\ {}_q\oplus\ \{x := x+1\}\}$$

With a standard invariant $I = \top$, the following variant V, together with $p(\cdot)$ and $d(\cdot)$ witness the AST of the loop :

$$V = x, \qquad \text{for} \quad d(v) = 1 \quad \text{and} \quad p(v) = \frac{1}{v+1}.$$

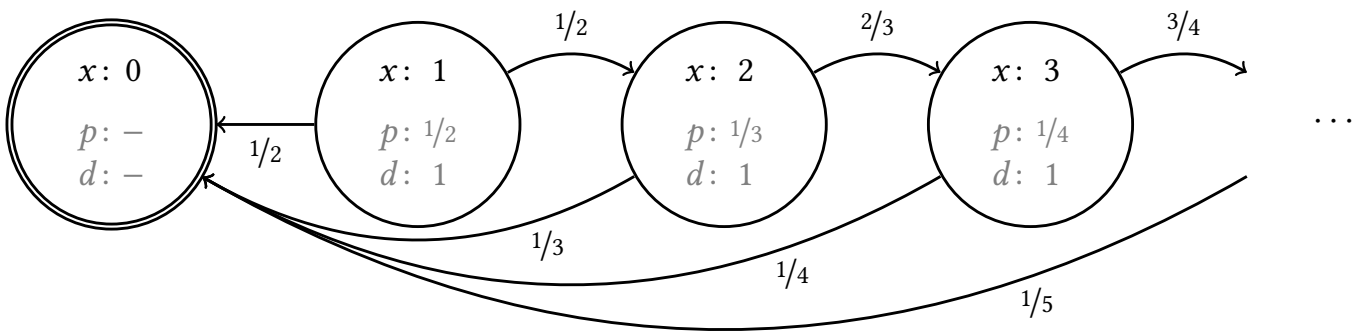


Figure – Execution of the escaping spline loop.

SOS Encoding

Given a loop $\texttt{while}(G)\{Com\}$ associated with a standard invariant I.

SOS Encoding

Given a loop $\texttt{while}(G)\{Com\}$ associated with a standard invariant I.

1. Assume template polynomials $V \in \mathbb{R}[\mathbf{x}]$ and $\bar{p}, d \in \mathbb{R}[x]$, with $\bar{p} = 1/p$.

SOS Encoding

Given a loop $\texttt{while}(G)\{Com\}$ associated with a standard invariant I.

1. Assume template polynomials $V \in \mathbb{R}[\mathbf{x}]$ and $\bar{p}, d \in \mathbb{R}[x]$, with $\bar{p} = 1/p$.
2. Determine template parameters in $V, \bar{p}, d$ s.t.

SOS Encoding

Given a loop $\texttt{while}(G)\{Com\}$ associated with a standard invariant I.

1. Assume template polynomials $V \in \mathbb{R}[\mathbf{x}]$ and $\bar{p}, d \in \mathbb{R}[x]$, with $\bar{p} = 1/p$.
2. Determine template parameters in $V, \bar{p}, d$ s.t.

Non-nagativity :

$$G \wedge I \implies V(\mathbf{x}) \geq 0 \quad \text{and} \quad x \geq 0 \implies \bar{p}(x) \geq 1 \quad \text{and} \quad x \geq 0 \implies d(x) \geq 0$$

Antitonicity :

$$x > 0 \wedge x \leq y \implies \bar{p}(x) \leq \bar{p}(y) \quad \text{and} \quad x > 0 \wedge x \leq y \implies d(y) \leq d(x)$$

Supermartingale :

$$V(\mathbf{x}) \geq [G \wedge I] \cdot \mathrm{awp}.Com.V(\mathbf{x})$$

Progress :

$$r > 0 \implies [G \wedge I \wedge V(\mathbf{x}) = r] \leq \bar{p}(r) \cdot \mathrm{wp}.Com.[V(\mathbf{x}) \leq r - d(r)]$$

SOS Encoding

Given a loop $\texttt{while}(G)\{Com\}$ associated with a standard invariant I.

1. Assume template polynomials $V \in \mathbb{R}[\mathbf{x}]$ and $\bar{p}, d \in \mathbb{R}[x]$, with $\bar{p} = 1/p$.
2. Determine template parameters in $V, \bar{p}, d$ s.t.

Non-nagativity :

$$G \wedge I \implies V(\mathbf{x}) \geq 0 \quad \text{and} \quad x \geq 0 \implies \bar{p}(x) \geq 1 \quad \text{and} \quad x \geq 0 \implies d(x) \geq 0$$

Antitonicity :

$$x > 0 \wedge x \leq y \implies \bar{p}(x) \leq \bar{p}(y) \quad \text{and} \quad x > 0 \wedge x \leq y \implies d(y) \leq d(x)$$

Supermartingale :

$$V(\mathbf{x}) \geq [G \wedge I] \cdot \mathrm{awp}.Com.V(\mathbf{x})$$

Progress-I :

$$r > 0 \wedge G \wedge I \wedge V(\mathbf{x}) = r \implies 1 \leq \bar{p}(r) \cdot \mathrm{wp}.Com.[V(\mathbf{x}) \leq r - d(r)]$$

Progress-II :

$$r > 0 \wedge \neg(G \wedge I \wedge V(\mathbf{x}) = r) \implies 0 \leq \bar{p}(r) \cdot \mathrm{wp}.Com.[V(\mathbf{x}) \leq r - d(r)]$$

Challenge

Progress-I :

$$r > 0 \wedge G \wedge I \wedge V(\mathbf{x}) = r \implies 1 \leq \bar{p}(r) \cdot \mathrm{wp}.Com.[V(\mathbf{x}) \leq r - d(r)]$$

Challenge

Progress-I :

$$r > 0 \wedge G \wedge I \wedge V(\mathbf{x}) = r \implies 1 \leq \bar{p}(r) \cdot \mathrm{wp}.Com.[V(\mathbf{x}) \leq r - d(r)]$$

3. Applying SOS relaxation on **Progress-I** yields that

$$\bar{p}(r) \cdot \mathrm{wp}.Com.[V(\mathbf{x}) \leq r - d(r)] - 1 - u_1 r - \ldots - v \cdot (V(\mathbf{x}) - r)$$

is an SOS for some $u_i \in \mathbb{R}_{\geq 0}$ and $v \in \mathbb{R}$.

Challenge

Progress-I :

$$r > 0 \wedge G \wedge I \wedge V(\mathbf{x}) = r \implies 1 \leq \bar{p}(r) \cdot \mathrm{wp}.Com.[V(\mathbf{x}) \leq r - d(r)]$$

3. Applying SOS relaxation on **Progress-I** yields that

$$\bar{p}(r) \cdot \mathrm{wp}.Com.[V(\mathbf{x}) \leq r - d(r)] - 1 - u_1 r - \ldots - v \cdot (V(\mathbf{x}) - r)$$

is an SOS for some $u_i \in \mathbb{R}_{\geq 0}$ and $v \in \mathbb{R}$.

Nonlinearity in parameters $\implies$ Cannot be relaxed to an SDP.

Potential Solution

Inspired by [Wang et al., 2019] : Assume a known compact *I*. Suppose that the parameter vector $\mathbf{a}$ of $V(\mathbf{a}, \mathbf{x})$ is represented as $\phi(\mathbf{a}) \leq 0$. Then *Lasserre's hierarchy of SOS programmings* [Lasserre, 2001, 2010] can be used to identify (through SDP solving) an inner-approximation of the feasible space of $\mathbf{a}$, i.e., $\phi'(\mathbf{a}) \leq 0$ with $\phi'(\mathbf{a}) \geq \phi(\mathbf{a})$. We can hence replace the **Progress-I** condition with a convex subset of $\phi'(\mathbf{a}) \leq 0$ and transform the whole problem to an SDP using SOS relaxation.

Potential Solution

Inspired by [Wang et al., 2019] : Assume a known compact *I*. Suppose that the parameter vector $\mathbf{a}$ of $V(\mathbf{a}, \mathbf{x})$ is represented as $\phi(\mathbf{a}) \leq 0$. Then *Lasserre's hierarchy of SOS programmings* [Lasserre, 2001, 2010] can be used to identify (through SDP solving) an inner-approximation of the feasible space of $\mathbf{a}$, i.e., $\phi'(\mathbf{a}) \leq 0$ with $\phi'(\mathbf{a}) \geq \phi(\mathbf{a})$. We can hence replace the **Progress-I** condition with a convex subset of $\phi'(\mathbf{a}) \leq 0$ and transform the whole problem to an SDP using SOS relaxation.

Assume a template for *I*? $\Longrightarrow$ Synthesizing invariant & variant, simultaneously!

Quantitative Craig Interpolants as Loop Invariants

Craig Interpolation

Definition (Interpolant [Craig, 1957])

Given ϕ and ψ in a theory $\mathcal{T}$ s.t. $\phi \wedge \psi \models_{\mathcal{T}} \bot$, a formula I is a *(reverse) interpolant* of ϕ and ψ if (1) $\phi \models_{\mathcal{T}} I$; (2) $I \wedge \psi \models_{\mathcal{T}} \bot$; and (3) $\mathit{var}(I) \subseteq \mathit{var}(\phi) \cap \mathit{var}(\psi)$.

Craig Interpolation

Definition (Interpolant [Craig, 1957])

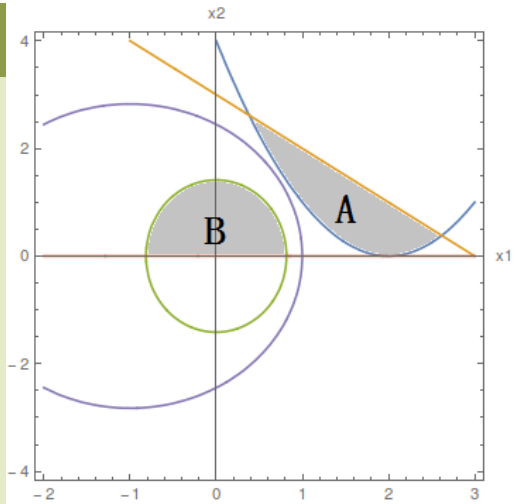
Given ϕ and ψ in a theory $\mathcal{T}$ s.t. $\phi \wedge \psi \models_{\mathcal{T}} \bot$, a formula I is a *(reverse) interpolant* of ϕ and ψ if (1) $\phi \models_{\mathcal{T}} I$; (2) $I \wedge \psi \models_{\mathcal{T}} \bot$; and (3) $\mathit{var}(I) \subseteq \mathit{var}(\phi) \cap \mathit{var}(\psi)$.

Example (over nonlinear $\mathcal{T}$)

$$A \mathrel{\widehat{=}} -x_1{}^2 + 4x_1 + x_2 - 4 \geq 0 \wedge -x_1 - x_2 + 3 - y^2 > 0$$

$$B \mathrel{\widehat{=}} -3x_1{}^2 - x_2{}^2 + 1 \geq 0 \wedge x_2 - z^2 \geq 0$$

$$I \mathrel{\widehat{=}} -3 + 2x_1 + x_1{}^2 + \frac{1}{2}x_2{}^2 > 0$$



Craig Interpolation

Definition (Interpolant [Craig, 1957])

Given ϕ and ψ in a theory $\mathcal{T}$ s.t. $\phi \wedge \psi \models_{\mathcal{T}} \bot$, a formula I is a *(reverse) interpolant* of ϕ and ψ if (1) $\phi \models_{\mathcal{T}} I$; (2) $I \wedge \psi \models_{\mathcal{T}} \bot$; and (3) $\mathit{var}(I) \subseteq \mathit{var}(\phi) \cap \mathit{var}(\psi)$.

Definition (Quantitative Craig Interpolant?)

Let $(\mathbb{E}, \sqsubseteq)$ be a complete lattice. Given $f, g \in \mathbb{E}$ s.t. $f \cdot g \sqsubseteq \lambda s.0$, an expectation $i \in \mathbb{E}$ is a *q-interpolant* of f and g if (1') $f \sqsubseteq i$; (2') $i \cdot g \sqsubseteq \lambda s.0$; and (3') $\mathit{var}(i) \subseteq \mathit{var}(f) \cap \mathit{var}(g)$.

Lehrstuhl für Informatik 2 Softwaremodellierung und Verifikation | RWTH AACHEN UNIVERSITY

Craig Interpolation

Definition (Interpolant [Craig, 1957])

Given ϕ and ψ in a theory $\mathcal{T}$ s.t. $\phi \wedge \psi \models_{\mathcal{T}} \bot$, a formula I is a *(reverse) interpolant* of ϕ and ψ if (1) $\phi \models_{\mathcal{T}} I$; (2) $I \wedge \psi \models_{\mathcal{T}} \bot$; and (3) $\mathit{var}(I) \subseteq \mathit{var}(\phi) \cap \mathit{var}(\psi)$.

Definition (Quantitative Craig Interpolant?)

Let $(\mathbb{E}, \sqsubseteq)$ be a complete lattice. Given $f, g \in \mathbb{E}$ s.t. $f \cdot g \sqsubseteq \lambda s.0$, an expectation $i \in \mathbb{E}$ is a *q-interpolant* of f and g if (1') $f \sqsubseteq i$; (2') $i \cdot g \sqsubseteq \lambda s.0$; and (3') $\mathit{var}(i) \subseteq \mathit{var}(f) \cap \mathit{var}(g)$.

Q1 : What is the intuitive meaning of (2')?

Q1' : What is the "negation" of an expectation?

Q2 : (3') may not hold for many examples?

Q2' : How to extract f and g from a probabilistic program?

Lehrstuhl für Informatik 2 Softwaremodellierung und Verifikation | RWTH AACHEN UNIVERSITY

Craig Interpolation

Definition (Interpolant [Craig, 1957])

Given ϕ and ψ in a theory $\mathcal{T}$ s.t. $\phi \wedge \psi \models_{\mathcal{T}} \bot$, a formula I is a *(reverse) interpolant* of ϕ and ψ if (1) $\phi \models_{\mathcal{T}} I$; (2) $I \wedge \psi \models_{\mathcal{T}} \bot$; and (3) $\mathit{var}(I) \subseteq \mathit{var}(\phi) \cap \mathit{var}(\psi)$.

Definition (Quantitative Craig Interpolant?)

Let $(\mathbb{E}, \sqsubseteq)$ be a complete lattice. Given $f, g \in \mathbb{E}$ s.t. $f \cdot g \sqsubseteq \lambda s.0$, an expectation $i \in \mathbb{E}$ is a *q-interpolant* of f and g if (1') $f \sqsubseteq i$; (2') $i \cdot g \sqsubseteq \lambda s.0$; and (3') $\mathit{var}(i) \subseteq \mathit{var}(f) \cap \mathit{var}(g)$.

- Q1 : What is the intuitive meaning of (2')?
- Q1' : What is the "negation" of an expectation?
- Q2 : (3') may not hold for many examples?
- Q2' : How to extract f and g from a probabilistic program?

Start from the special case where f, g are of the form $[P]$, with $P \in \mathbb{P}$.

Lehrstuhl für Informatik 2 Softwaremodellierung und Verifikation | RWTH AACHEN UNIVERSITY

Summary