Chapter 21
Virtual Substitution & Real Arithmetic

Synopsis This chapter advances the understanding of real arithmetic by generalizing the ideas from the previous chapter to linear and quadratic *inequalities*. As in the previous chapter, the main workhorse will again be virtual substitutions that pretend to substitute a generalized expression into a logical formula by equivalently rephrasing each occurrence. The required virtual substitutions will, however, go beyond square root substitutions but cover infinities and infinitesimals, instead, in order to capture the fact that inequalities can also be satisfied without satisfying equality.

21.1 Introduction

Reasoning about cyber-physical systems and hybrid systems requires understanding and handling their real arithmetic, which can be challenging, because cyber-physical systems can have complex behavior. Differential dynamic logic and its proof calculus [6–8] reduce the verification of hybrid systems to real arithmetic. How arithmetic interfaces with proofs has already been discussed in Chap. 6. How real arithmetic with linear and quadratic equations can be handled by virtual substitution has been shown in Chap. 20. This chapter shows how virtual substitution for quantifier elimination in real arithmetic extends to the case of linear and quadratic inequalities.

The results in this chapter are based on the literature [13]. The chapter adds substantial intuition and motivation that is helpful for following the technical development. More information about virtual substitution can be found in the literature [13]. See, e.g., [1, 2, 5, 9] for an overview of other techniques for real arithmetic.

The most important learning goals of this chapter are:

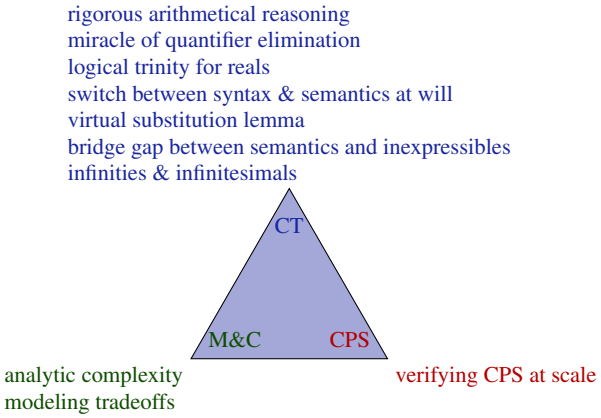
Modeling and Control: This chapter refines the indirect impact that the previous chapter had on CPS models and controls by informing the reader about the consequences of the analytic complexity resulting from different arithmetical modeling tradeoffs. There are subtle analytic consequences from different arithmetic formulations of similar questions that can have an impact on finding the right tradeoffs for expressing a CPS model. In practical terms, a safe distance

A. Platzer, *Logical Foundations of Cyber-Physical Systems*,
https://doi.org/10.1007/978-3-319-63588-0_21

of car x to a stop light m could equally well be captured as $x \leq m$ or as $x < m$, for example, if only we knew the impact of this decision on the resulting real arithmetic.

Computational Thinking: The primary purpose of this chapter is to understand how arithmetical reasoning, which is crucial for CPS, can be done rigorously and automatically not just for the equations considered in Chap. 20 but also for inequalities. While formulas involving sufficiently many quadratic equations among other inequalities can be handled with the techniques from Chap. 20, such extensions are crucial for proving arithmetic formulas that involve only inequalities, which happens rather frequently in the world of CPS, where many questions concern inequality bounds on distances. Developing an intuition for the working principles of real-arithmetic decision procedures can be very helpful for developing strategies to verify CPS models at scale. We will again see the conceptually very important device of the logical trinity: the flexibility of moving back and forth between syntax and semantics at will. Virtual substitutions will again allow us to move back and forth at will between syntax and semantics. This time, however, square roots will not be all there is to it, but the logical trinity will lead us to ideas from nonstandard analysis to bridge the gap to semantic operations that are inexpressible otherwise in first-order logic of real arithmetic.

CPS Skills: This chapter has an indirect impact on CPS skills, because it discusses useful pragmatics of CPS analysis for modeling and analysis tradeoffs that enable CPS verification at scale.



21.2 Recap: Square Root $\sqrt{\cdot}$ Virtual Substitutions for Quadratics

Recall the way to handle quantifier elimination for linear or quadratic equations from Chap. 20 by virtually substituting in its symbolic solutions $x = -c/b$ or $x = (-b \pm \sqrt{b^2 - 4ac})/(2a)$, respectively

Theorem 20.2 (Virtual substitution of quadratic equations). *For quantifier-free formula F with $x \notin FV(a), FV(b), FV(c)$, the following equivalence is valid over $\mathbb{R}$:*

$$
\begin{aligned}
&a \neq 0 \vee b \neq 0 \vee c \neq 0 \rightarrow \\
&\quad \Big(\exists x\,(ax^2 + bx + c = 0 \wedge F) \leftrightarrow \\
&\qquad a = 0 \wedge b \neq 0 \wedge F_x^{-c/b} \\
&\qquad \vee\, a \neq 0 \wedge b^2 - 4ac \geq 0 \wedge \big(F_x^{(-b+\sqrt{b^2-4ac})/(2a)} \vee F_x^{(-b-\sqrt{b^2-4ac})/(2a)}\big)\Big)
\end{aligned}
$$

When using virtual substitutions of square roots from Chap. 20, the resulting formula on the right-hand side of the bi-implication is quantifier-free and can be chosen for $\mathrm{QE}(\exists x\,(ax^2 + bx + c = 0 \wedge F))$ as long as it is not the case that $a = b = c = 0$. In case $a = b = c = 0$, another formula in F needs to be considered for directing quantifier elimination by commuting and reassociating $\wedge$, because the equation $ax^2 + bx + c = 0$ is noninformative if $a = b = c = 0$, e.g., when a, b, c are the zero polynomial or even if they just have a common root.

The equivalent formula on the right-hand side of the bi-implication in Theorem 20.2 is a formula in the first-order logic of real arithmetic when using the virtual substitution of square root expressions defined in Chap. 20.

21.3 Infinity ∞ Virtual Substitution

Theorem 20.2 addresses the case where the quantified variable occurs in a linear or quadratic equation, in which case it is efficient to use Theorem 20.2, because there are at most three symbolic points to consider corresponding to the respective solutions of the equation. But what do we do if the quantified variable only occurs in inequalities? Then Theorem 20.2 does not help the slightest bit. Consider a formula of the form

$$\exists x\,(ax^2 + bx + c \leq 0 \wedge F) \qquad (x \notin \mathrm{FV}(a), \mathrm{FV}(b), \mathrm{FV}(c)) \tag{21.1}$$

where x does not occur in a, b, c. Under the conditions from Theorem 20.2, the possible solutions $-c/b, (-b + \sqrt{d})/(2a), (-b - \sqrt{d})/(2a)$ from Theorem 20.2 continue to be options for solutions of (21.1), because one way of satisfying the weak inequality $ax^2 + bx + c \leq 0$ is by satisfying the equation $ax^2 + bx + c = 0$. So if F is

true for any of those solutions of the quadratic equation (under the auspices of the additional constraints on a, b, c), then (21.1) holds as well.

Yet, even if those points do not work out, the weak inequality in (21.1) allows for more possible solutions than the equation does. For example, if $a = 0, b > 0$, then sufficiently small values of x would satisfy $0x^2 + bx + c \leq 0$. Also, if $a < 0$, then sufficiently small values of x would satisfy $ax^2 + bx + c \leq 0$, because x^2 grows faster than x and, thus the negative ax^2 ultimately overcomes any contribution of bx and c to the value of $ax^2 + bx + c$. But if we literally substituted each such smaller value of x into F, that would quickly diverge into the full substitution $\bigvee_{t \in T} F_x^t$ for the uninsightful case of all real numbers $T \stackrel{\text{def}}{=} \mathbb{R}$ from Chap. 20. So we have to be more clever than that.

Now, one possibile way of pursuing this line of thought may be to substitute smaller and smaller values for x into (21.1) and see if one of those happens to work. There is a much better way though. The only really small value that has to be substituted into (21.1) for x to see whether it happens to work is one that is so negative that it is smaller than all others: $-\infty$, which is the lower limit of all negative real numbers. Alternatively, $-\infty$ can be understood as being "always as negative as needed, i.e., more negative than anything else." Think of $-\infty$ as being built out of elastic rubber so that it always ends up being smaller when compared to any actual real number, because the elastic number $-\infty$ simply shrinks every time it is compared to any other number. Analogously, ∞ is the upper limit of all real numbers or "always as positive as needed, i.e., more positive than anything else." The elastic rubber version of understanding ∞ is such that ∞ always grows as needed every time it is compared to any other number.

Let $\infty, -\infty$ be *positive and negative infinities*, respectively, i.e., choose extra elements $\infty, -\infty \notin \mathbb{R}$ with $-\infty < r < \infty$ for all $r \in \mathbb{R}$. Formulas of real arithmetic can be substituted with $\pm\infty$ for a variable x in the compactified reals $\mathbb{R} \cup \{\infty, -\infty\}$. Yet, just like with square root expressions, $\pm\infty$ do not actually need to ever truly occur in the resulting formula, because substitution of infinities into formulas can be defined differently. For example, $(x + 5 > 0)_x^{-\infty}$ will be *false*, while $(x + 5 < 0)_x^{-\infty}$ is *true*.

Definition 21.1 (Infinite virtual substitution). *Substitution of the infinity* $-\infty$ for x into an atomic formula for a polynomial $p \stackrel{\text{def}}{=} \sum_{i=0}^{n} a_i x^i$ with polynomials a_i that do not contain x is defined by the following equivalences:

$$(p = 0)_{\vec{x}}^{-\infty} \equiv \bigwedge_{i=0}^{n} a_i = 0 \tag{21.2}$$

$$(p \leq 0)_{\vec{x}}^{-\infty} \equiv (p < 0)_{\vec{x}}^{-\infty} \vee (p = 0)_{\vec{x}}^{-\infty} \tag{21.3}$$

$$(p < 0)_{\vec{x}}^{-\infty} \equiv p(-\infty) < 0 \tag{21.4}$$

$$(p \neq 0)_{\vec{x}}^{-\infty} \equiv \bigvee_{i=0}^{n} a_i \neq 0 \tag{21.5}$$

Lines (21.2) and its dual (21.5) use that the only equation of real arithmetic that infinities $\pm\infty$ satisfy is the trivial equation $0 = 0$. Line (21.3) uses the equivalence $p \leq 0 \equiv p < 0 \vee p = 0$ and is equal to $(p < 0 \vee p = 0)^{-\infty}_{\tilde{x}}$ by the substitution base from Sect. 20.3.2. Line (21.4) uses a simple inductive definition based on the *degree*, $\deg(p)$, the highest power of the variable x in the polynomial p, to characterize whether p is ultimately negative at $-\infty$ (or sufficiently negative numbers):

Let $p \stackrel{\text{def}}{=} \sum_{i=0}^{n} a_i x^i$ with polynomials a_i that do not contain x. Whether p is ultimately negative at $-\infty$, suggestively written $p(-\infty) < 0$, is easy to characterize by induction on the degree of the polynomial:

$$p(-\infty) < 0 \stackrel{\text{def}}{\equiv} \begin{cases} p < 0 & \text{if } \deg(p) \leq 0 \\ (-1)^n a_n < 0 \vee (a_n = 0 \wedge (\sum_{i=0}^{n-1} a_i x^i)(-\infty) < 0) & \text{if } \deg(p) > 0 \end{cases}$$

$p(-\infty) < 0$ is true in a state in which $\lim_{x \to -\infty} p(x) < 0$.

The first line captures that the sign of polynomials of degree 0 in the variable x does not depend on x, so $p(-\infty) < 0$ iff the polynomial of degree 0 in x is negative (which may still depend on the value of other variables in $p = a_0$ but not on x). The second line captures that the sign at $-\infty$ of a polynomial of degree $n = \deg(p) > 0$ is determined by the degree-modulated sign of its leading coefficient a_n, because for x of sufficiently big absolute value, the value of $a_n x^n$ will dominate all lower-degree values, whatever their coefficients are. For even $n > 0$, $x^n > 0$ while $x^n < 0$ for odd n at $-\infty$. In case the leading coefficient a_n evaluates to zero, the value of p at $-\infty$ depends on the value at $-\infty$ of the remaining polynomial $\sum_{i=0}^{n-1} a_i x^i$ of lower degree, which can be determined recursively as $(\sum_{i=0}^{n-1} a_i x^i)(-\infty) < 0$. Note that the degree of the 0 polynomial is sometimes considered to be $-\infty$, which explains why $\deg(p) \leq 0$ is used in line 1 instead of $\deg(p) = 0$.

Substitution of ∞ for x into an atomic formula can be defined similarly, except that the sign factor $(-1)^n$ disappears, because $x^n > 0$ at ∞ whatever value $n > 0$ has. Substitution of ∞ or of $-\infty$ for x into other first-order formulas is then defined on this basis as in Sect. 20.3.2.

Example 21.1 (Sign of quadratic polynomials at $-\infty$). Using this principle to check systematically under which circumstances the quadratic inequality from (21.1) evaluates to *true* yields the answer from our earlier ad-hoc analysis of what happens for sufficiently small values of x:

$$\begin{aligned} (ax^2 + bx + c < 0)^{-\infty}_{\tilde{x}} &\equiv (-1)^2 a < 0 \vee a = 0 \wedge ((-1)b < 0 \vee b = 0 \wedge c < 0) \\ &\equiv a < 0 \vee a = 0 \wedge (b > 0 \vee b = 0 \wedge c < 0) \\ (ax^2 + bx + c \leq 0)^{-\infty}_{\tilde{x}} &\equiv (ax^2 + bx + c < 0)^{-\infty}_{\tilde{x}} \vee a = b = c = 0 \\ &\equiv a < 0 \vee a = 0 \wedge (b > 0 \vee b = 0 \wedge c < 0) \vee a = b = c = 0 \end{aligned}$$

One representative example for each of those disjuncts is illustrated in Fig. 21.1. In the same way, the virtual substitution can be used to see under which circumstances

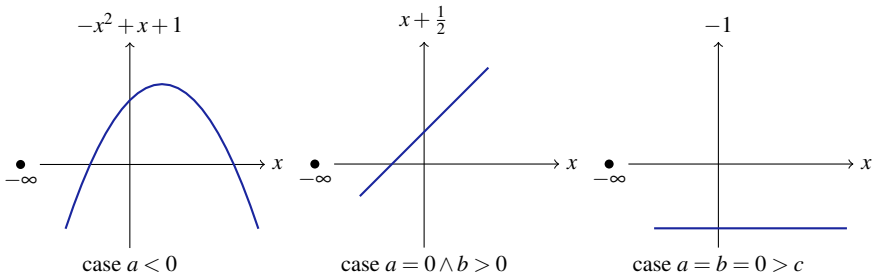


Fig. 21.1 Illustration of the value of different quadratic functions p where $p_{\overline{x}}^{-\infty} \equiv \mathit{true}$

the remainder formula F from (21.1) also evaluates to *true* for sufficiently small values of x, which is the case exactly when $F_{\overline{x}}^{-\infty}$ evaluates to *true*.

The crucial thing to note is again that the *virtual substitution* of infinities $\pm\infty$ for x in F giving $F_{\overline{x}}^{\pm\infty}$ from Definition 21.1 is semantically equivalent to the result $F_x^{\pm\infty}$ of the literal substitution replacing x with $\pm\infty$, but operationally different, because the virtual substitution never introduces actual infinities so remains in proper $\mathrm{FOL}_{\mathbb{R}}$.

Lemma 21.1 (Virtual substitution lemma for infinities). *The result $F_{\overline{x}}^{-\infty}$ of the virtual substitution is semantically equivalent to the result $F_x^{-\infty}$ of the literal substitution. A language extension yields this validity:*

$$F_x^{-\infty} \leftrightarrow F_{\overline{x}}^{-\infty}$$

Keep in mind that the result $F_{\overline{x}}^{-\infty}$ of virtual substitution is a proper formula of $\mathrm{FOL}_{\mathbb{R}}$, while the literal substitution $F_x^{-\infty}$ can only be considered a formula in an extended logic such as $\mathrm{FOL}_{\mathbb{R}\cup\{-\infty,\infty\}}$ that allows for infinite quantities. The same property holds for $F_{\overline{x}}^{\infty}$.

Note that the situation is, in a sense, the converse of Lemma 20.2, where the square root expressions were already in the semantic domain $\mathbb{R}$, and just had to be made accessible in the syntactic formulas via virtual substitutions. In Lemma 21.1, instead, virtual substitutions already know more about infinities $\pm\infty$ than the semantic domain $\mathbb{R}$ does, which is why the semantic domain needs an extension to $\mathbb{R}\cup\{-\infty,\infty\}$ for the alignment in Lemma 21.1.

21.4 Infinitesimal ε Virtual Substitution

Theorem 20.2 addresses the case where the quantified variable occurs in a linear or quadratic equation and the virtual substitution in Sect. 21.3 adds the case of sufficiently small values of x to handle $ax^2+bx+c\leq 0$. Consider a formula of the form

$$\exists x\,(ax^2+bx+c<0\wedge F) \qquad (x\notin \mathrm{FV}(a),\mathrm{FV}(b),\mathrm{FV}(c)) \tag{21.7}$$

Expedition 21.1 (Infinite challenges with infinities in extended reals)

The set $\mathbb{R} \cup \{-\infty, \infty\}$ is seemingly easily written down as a semantic domain of extended reals. What exactly do we mean by it, though? We mean the set of reals to which we adjoin two new elements, denoted $-\infty$ and ∞, which are the minimum and maximum elements of the ordering $\leq$:

$$\forall x\,(-\infty \leq x \leq \infty) \tag{21.6}$$

This turns $\mathbb{R} \cup \{-\infty, \infty\}$ into a complete lattice, because every subset has a supremum and an infimum. The extended reals are a compactification of $\mathbb{R}$. But where does that leave the other arithmetic properties of $\mathbb{R}$? What is $\infty + 1$ or $\infty + x$ when ∞ is already infinitely big? The compatibility of $\leq$ with $+$ expects $\infty \leq \infty + x$ at least for all $x \geq 0$. By (21.6) also $\infty + x \leq \infty$. Because ∞ is so infinitely big, the same $\infty + x = \infty$ is expected even for all x, except $-\infty$. The compatibility of $\leq$ with $\cdot$ expects $\infty \leq \infty \cdot x$ at least for all $x \geq 1$. By (21.6) also $\infty \cdot x \leq \infty$. Since ∞ is infinitely big, the same $\infty \cdot x = \infty$ is expected even for all $x > 0$:

$$\begin{aligned}
\infty + x &= \infty && \text{for all } x \neq -\infty\\
-\infty + x &= -\infty && \text{for all } x \neq \infty\\
\infty \cdot x &= \infty && \text{for all } x > 0\\
\infty \cdot x &= -\infty && \text{for all } x < 0\\
-\infty \cdot x &= -\infty && \text{for all } x > 0\\
-\infty \cdot x &= \infty && \text{for all } x < 0
\end{aligned}$$

This extension sounds reasonable. But the resulting set $\mathbb{R} \cup \{-\infty, \infty\}$ is not a field! Otherwise ∞ would have an additive inverse. But what x would satisfy $\infty + x = 0$? One might guess $x = -\infty$, but then one would also expect $0 = \infty + (-\infty) = \infty + (-\infty + 1) = (\infty + (-\infty)) + 1 = 0 + 1 = 1$, which is not a good idea to adopt for proving anything at all in a sound way. Instead, problematic terms remain explicitly undefined:

$$\begin{aligned}
\infty - \infty &= \text{undefined}\\
0 \cdot \infty &= \text{undefined}\\
\pm\infty / \pm\infty &= \text{undefined}\\
1/0 &= \text{undefined}
\end{aligned}$$

Since these conventions make infinities somewhat subtle, we happily remember that the only thing we need them for is to make sense of inserting sufficiently negative (or sufficiently positive) numbers into inequalities to satisfy them. That is still mostly harmless.

In this case, the roots from Theorem 20.2 will not help, because they satisfy the equation $ax^2+bx+c=0$ but not the strict inequality $ax^2+bx+c<0$. The virtual substitution of $-\infty$ for x from Sect. 21.3 still makes sense to consider, because the arbitrarily small negative numbers that it corresponds to might indeed satisfy F and $ax^2+bx+c<0$. If $-\infty$ does not work, however, the solution of (21.7) might be *near* one of the roots of $ax^2+bx+c=0$, just *slightly off* so that $ax^2+bx+c<0$ is actually satisfied rather than the equation $ax^2+bx+c=0$. How far off? Well, saying that exactly is again difficult, because any particular real number might already be too large in absolute value, depending on the constraints in the remainder of F. Again, this calls for quantities that are always as small as we need them to be.

Sect. 21.3 used a negative quantity that is so small that it is smaller than all negative numbers and hence infinitely small (but infinitely large in absolute value). The negative infinity $-\infty$ is smaller no matter what other number we compare it with. Analyzing (21.7) needs *positive* quantities that are infinitely small and hence also infinitely small in absolute value. Infinitesimals are positive quantities that are always smaller than all positive real numbers, i.e., "always as small as needed." Think of them as built out of elastic rubber so that they always shrink as needed when compared with any actual positive real number so that the infinitesimals end up being smaller than positive reals. Of course, the infinitesimals are much bigger than negative numbers. Another way of looking at infinitesimals is that they are the multiplicative inverses of $\pm\infty$.

A *positive infinitesimal* ε is positive ($\infty > \varepsilon > 0$) and an extended real that is *infinitesimal*, i.e., positive but smaller than all positive real numbers ($\varepsilon < r$ for all $r \in \mathbb{R}$ with $r > 0$).

Note 83 (Infinitesimals in polynomials) All nonzero univariate polynomials $p \in \mathbb{R}[x]$ with real coefficients satisfy the following cases infinitesimally near any real point $\zeta \in \mathbb{R}$:

1. $p(\zeta+\varepsilon) \neq 0$
   That is, infinitesimals ε are always so small that they never yield roots of any equation, except the trivial zero polynomial. Whenever it looks like there might be a root, the infinitesimal just becomes a bit smaller to avoid satisfying the equation. Nonzero univariate polynomials $p(x)$ only have finitely many roots, so the infinitesimals will take care to avoid all of them by becoming just a little smaller.
2. If $p(\zeta) \neq 0$ then $p(\zeta)p(\zeta+\varepsilon) > 0$.
   That is, p has constant sign on infinitesimal neighborhoods of nonroots ζ. If the neighborhood around ζ is small enough (and for an infinitesimal it will be), then the polynomial will not change sign on that interval, because the sign will only change after passing one of the roots.
3. $0=p(\zeta)=p'(\zeta)=p''(\zeta)=\ldots=p^{(k-1)}(\zeta)\neq p^{(k)}(\zeta)$ then $p^{(k)}(\zeta)p(\zeta+\varepsilon)>0$.
   That is the first nonzero derivative of p at ζ determines the sign of p in small enough neighborhoods of ζ (infinitesimal neighborhoods will be small enough), because the sign only changes after passing a root.

Definition 21.2 (Infinitesimal virtual substitution). *Substitution of an infinitesimal expression* $e+\varepsilon$ with a square root expression $e=(a+b\sqrt{c})/d$ and a positive infinitesimal ε for x into a polynomial $p=\sum_{i=0}^{n} a_i x^i$ with polynomials a_i that do not contain x is defined by the following equivalences:

$$(p=0)_{\tilde{x}}^{e+\varepsilon} \equiv \bigwedge_{i=0}^{n} a_i = 0 \tag{21.8}$$

$$(p\leq 0)_{\tilde{x}}^{e+\varepsilon} \equiv (p<0)_{\tilde{x}}^{e+\varepsilon} \vee (p=0)_{\tilde{x}}^{e+\varepsilon} \tag{21.9}$$

$$(p<0)_{\tilde{x}}^{e+\varepsilon} \equiv (p^+<0)_{\tilde{x}}^{e} \tag{21.10}$$

$$(p\neq 0)_{\tilde{x}}^{e+\varepsilon} \equiv \bigvee_{i=0}^{n} a_i \neq 0 \tag{21.11}$$

Lines (21.8) and its dual (21.11) use that infinitesimal offsets satisfy no equation except the trivial equation 0=0 (Case 1 of Note 83), which makes infinitesimals and infinities behave the same as far as equations go. Line (21.9) again uses the equivalence $p\leq 0 \equiv p<0\vee p=0$. Line (21.10) checks whether the sign of p at the square root expression e is already negative (which will make p inherit the same negative sign after an infinitesimal offset at $e+\varepsilon$ by Case 2) or will immediately become negative using a recursive formulation of immediately becoming negative that uses higher derivatives (which determine the sign by Case 3). The lifting to arbitrary quantifier-free formulas of real arithmetic is again by substitution into all atomic subformulas and equivalences such as $(p>q)\equiv(p-q>0)$ as defined in Chap. 20. Note that, for the case $(p<0)_{\tilde{x}}^{e+\varepsilon}$, the (non-infinitesimal) square root expression e gets virtually substituted in for x into a formula $p^+<0$, which characterizes whether p becomes negative at or immediately after x (which will be virtually substituted by the intended square root expression e momentarily).

Whether p is immediately negative at x, i.e., negative itself or 0 and with a derivative p' that makes it negative on an infinitesimal interval $(x,x+\varepsilon]$, suggestively written $p^+<0$, can be characterized recursively:

$$p^+<0 \stackrel{\text{def}}{\equiv} \begin{cases} p<0 & \text{if } \deg(p)\leq 0 \\ p<0\vee(p=0\wedge(p')^+<0) & \text{if } \deg(p)>0 \end{cases}$$

$p^+<0$ is true in a state in which $\lim_{y\to x^+} p(x) = \lim_{y\searrow x} p(x) = \lim_{\substack{y>x\\ y\to x}} p(x) < 0$ holds for the limit of p at x from the right.

The first line captures that the sign of polynomials of degree 0 in the variable x does not depend on x, so they are negative at x iff the polynomial $p=a_0$ that has degree 0 in x is negative (which may still depend on the value of other variables in a_0). The second line captures that the sign at $x+\varepsilon$ of a non-constant polynomial is still negative if it is negative at x (because $x+\varepsilon$ is not far enough away from x for any

sign change by Case 2) or if x is a root of p but its derivative p' at x is immediately negative, since the first nonzero derivative at x determines the sign near x by Case 3.

Example 21.2 (Sign of quadratic polynomials after root). Using this principle to check under which circumstances the quadratic strict inequality from (21.7) evaluates to *true* at the point $(-b+\sqrt{b^2-4ac})/(2a)+\varepsilon$, i.e., right after its quadratic root $(-b+\sqrt{b^2-4ac})/(2a)$, leads to the following computation:

$$\begin{aligned}&(ax^2+bx+c)^+ < 0\\ \equiv\ & ax^2+bx+c<0 \vee ax^2+bx+c=0 \wedge (2ax+b<0 \vee 2ax+b=0 \wedge 2a<0)\end{aligned}$$

with successive derivatives to break ties (i.e., 0 signs in previous derivatives). Hence,

$$\begin{aligned}&(ax^2+bx+c<0)_{\bar{x}}^{(-b+\sqrt{b^2-4ac})/(2a)+\varepsilon} \equiv ((ax^2+bx+c)^+ < 0)_{\bar{x}}^{(-b+\sqrt{b^2-4ac})/(2a)} \equiv\\ &(ax^2+bx+c<0 \vee ax^2+bx+c=0 \wedge (2ax+b<0 \vee 2ax+b=0 \wedge 2a<0))_{\bar{x}}^{(-b+\sqrt{b^2-4ac})/(2a)}\\ \equiv\ & 0\cdot 1<0 \vee 0=0 \wedge (\underbrace{(0<0 \vee 4a^2\le 0 \wedge (0<0 \vee -4a^2(b^2-4ac)<0))}_{(2ax+b<0)_{\bar{x}}^{(-b+\sqrt{b^2-4ac})/(2a)}} \vee \underbrace{0=0}_{(2ax+b=0)_{\bar{x}}^{\cdots}} \wedge \underbrace{2a1<0)}_{(2a<0)_{\bar{x}}^{\cdots}}\\ &\equiv 4a^2\le 0 \wedge -4a^2(b^2-4ac)<0 \vee 2a<0\end{aligned}$$

because the square root virtual substitution of its own root $(-b+\sqrt{b^2-4ac})/(2a)$ into ax^2+bx+c gives $(ax^2+bx+c)_{\bar{x}}^{(-b+\sqrt{b^2-4ac})/(2a)}=0$ by construction (compare Example 20.5). The virtual substitution into another polynomial $2ax+b$ gives

$$\begin{aligned}(2ax+b)_{\bar{x}}^{(-b\pm\sqrt{b^2-4ac})/(2a)} &\equiv 2a\cdot(-b\pm\sqrt{b^2-4ac})/(2a)+b\\ &= (-2ab+\pm 2a\sqrt{b^2-4ac})/(2a)+b\\ &= (\cancel{-2ab}+\cancel{2ab}+\pm 2a\sqrt{b^2-4ac})/(2a)\\ &= (0+\pm 2a\sqrt{b^2-4ac})/(2a)\end{aligned}$$

The resulting formula can be further simplified internally to

$$\begin{aligned}(ax^2+bx+c<0)_{\bar{x}}^{(-b+\sqrt{b^2-4ac})/(2a)+\varepsilon} &\equiv 4a^2\le 0 \wedge -4a^2(b^2-4ac)<0 \vee 2a<0\\ &\equiv 2a<0\end{aligned}$$

because the first conjunct $4a^2\le 0\equiv a=0$ and, with $a=0$, the second conjunct simplifies to $-4a^2(b^2-4ac)_a^0=-0(b^2)<0$, which is impossible in the reals. This answer makes sense. Indeed, exactly if $2a<0$ will a quadratic polynomial still evaluate to $ax^2+bx+c<0$ right after its second root $(-b+\sqrt{b^2-4ac})/(2a)$. Fig. 21.2 illustrates how this relates to the parabola pointing downwards, because of $2a<0$.

Formulas such as this one ($2a<0$) are the result of a quantifier elimination procedure. If the formula after quantifier elimination is either *true* or *false*, then you know for sure that the formula is valid (*true*) or unsatisfiable (*false*), respectively.

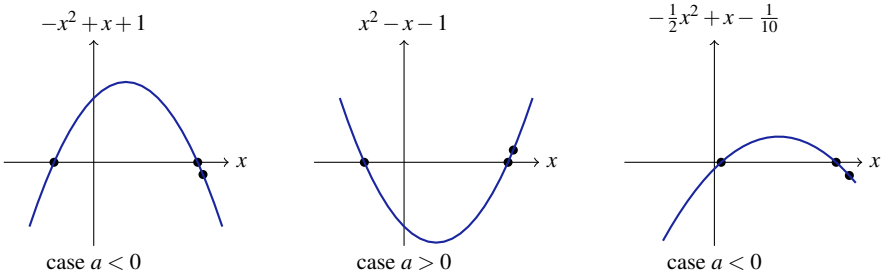


Fig. 21.2 Illustration of the sign after the second root for quadratic functions p

If the result of quantifier elimination is *true*, for example, KeYmaera X completes proof branches (marked by proof rule $\mathbb{R}$ in our sequent proofs). However, quantifier elimination can also return other formulas, such as $2a < 0$, which are equivalent to the formula to which quantifier elimination has been applied. In particular, they identify under exactly which circumstance that corresponding quantified formula is true. This can be very useful for identifying the missing assumptions to make a proof work and the corresponding statement true.

Note 84 (Quantifier elimination identifies requirements) If the outcome of quantifier elimination is the formula *true*, the corresponding formula is valid. If it is the formula *false*, the corresponding formula is not valid (and even unsatisfiable). In between, i.e., when quantifier elimination results in a logical formula that is sometimes false and sometimes true, then this formula identifies exactly the missing requirements that are needed to make the desired formula true. This can be useful to synthesize missing requirements. Take care, however, not to work with universal closures, in which case *true* and *false* are the only possible outcomes.

The crucial thing to note about the process is that the *virtual substitution* of infinitesimal expressions $e+\varepsilon$ for x in F giving $F^{e+\varepsilon}_{\overline{x}}$ from Definition 21.2 is semantically equivalent to the result $F^{e+\varepsilon}_{x}$ of the literal substitution replacing x with $e+\varepsilon$, but operationally different, because it never introduces actual infinitesimals.

Lemma 21.2 (Virtual substitution lemma for infinitesimals). *The result $F^{e+\varepsilon}_{\overline{x}}$ of the virtual substitution is semantically equivalent to the result $F^{e+\varepsilon}_{x}$ of the literal substitution. A language extension yields this validity:*

$$F^{e+\varepsilon}_{x} \leftrightarrow F^{e+\varepsilon}_{\overline{x}}$$

Keep in mind that the result $F^{e+\varepsilon}_{\overline{x}}$ of virtual substitution is a proper formula of $\text{FOL}_{\mathbb{R}}$, while the literal substitution $F^{e+\varepsilon}_{x}$ could only be considered a formula in an extended logic such as $\text{FOL}_{\mathbb{R}[\varepsilon]}$ that allows for infinitesimal quantities from nonstandard analysis. Computationally more efficient substitutions of infinitesimals have been reported elsewhere [3].

Expedition 21.2 (Nonstandard analysis: infinite challenges with infinitesimal ε)

Infinite quantities in the extended reals $\mathbb{R} \cup \{-\infty, \infty\}$ already needed some attention to stay away from undefined expressions. Infinitesimals are infinitely more subtle than infinities. Real numbers are Archimedean, i.e., for every nonzero $x \in \mathbb{R}$, there is an $n \in \mathbb{N}$ such that

$$|\underbrace{x + x + \cdots + x}_{n \text{ times}}| > 1$$

Infinitesimals are non-Archimedean, because it does not matter how often you add ε, it still won't sum to one. There is a myriad of ways of making sense of infinitesimal quantities in nonstandard analysis, including surreal numbers, superreal numbers, and hyperreals. In a sense, infinitesimal quantities can be considered to be multiplicative inverses of infinities, but bring up many subtleties. For example, if an infinitesimal ε is added to $\mathbb{R}$, then the following terms need to denote values and satisfy ordering relations:

$$\varepsilon^2 \quad \varepsilon \quad x^2 + \varepsilon \quad (x+\varepsilon)^2 \quad x^2 + 2\varepsilon x + 5\varepsilon + \varepsilon^2$$

Fortunately, a rather tame version of infinitesimals is enough for the context of virtual substitution. The crucial properties of infinitesimals we need are [4]:

$$\begin{aligned} &\varepsilon > 0 \\ &\forall x \in \mathbb{R}\,(x > 0 \to \varepsilon < x) \end{aligned}$$

That is, the infinitesimal ε is positive and smaller than all positive reals.

21.5 Quantifier Elimination by Virtual Substitution for Quadratics

The following quantifier elimination technique due to Weispfenning [13] works for formulas with a quantified variable that occurs at most quadratically.

Theorem 21.1 (Virtual substitution of quadratic constraints). *Let F be a quantifier-free formula in which all atomic formulas are of quadratic form $ax^2 + bx + c \sim 0$ for polynomials a, b, c that do not mention variable x (that is, $x \notin FV(a), FV(b), FV(c)$) with some comparison operator $\sim \in \{=, \leq, <, \neq\}$ and corresponding discriminant $d = b^2 - 4ac$. Then $\exists x F$ is equivalent over $\mathbb{R}$ to the following quantifier-free formula:*

$$
\begin{array}{l}
F_{\tilde{x}}^{-\infty} \\
\vee \bigvee_{ax^2+bx+c\left\{\begin{smallmatrix}=\\ \leq\end{smallmatrix}\right\}0\in F} \left(a{=}0 \wedge b{\neq}0 \wedge F_{\tilde{x}}^{-c/b} \vee a{\neq}0 \wedge d{\geq}0 \wedge (F_{\tilde{x}}^{(-b+\sqrt{d})/(2a)} \vee F_{\tilde{x}}^{(-b-\sqrt{d})/(2a)})\right) \\
\vee \bigvee_{ax^2+bx+c\left\{\begin{smallmatrix}\neq\\ <\end{smallmatrix}\right\}0\in F} \left(a{=}0 \wedge b{\neq}0 \wedge F_{\tilde{x}}^{-c/b+\varepsilon} \vee a{\neq}0 \wedge d{\geq}0 \wedge (F_{\tilde{x}}^{(-b+\sqrt{d})/(2a)+\varepsilon} \vee F_{\tilde{x}}^{(-b-\sqrt{d})/(2a)+\varepsilon})\right)
\end{array}
$$

Proof. The proof is an extended form of the proof reported in the literature [13]. The proof first considers the literal substitution of square root expressions, infinities, and infinitesimals and then, as a second step, uses that the virtual substitutions that avoid square root expressions, infinities, and infinitesimals are equivalent (Lemma 20.2, 21.1 and 21.2). Let G denote the quantifier-free right-hand side so that the validity of the following formula needs to be shown:

$$\exists x F \leftrightarrow G \tag{21.12}$$

The implication from the quantifier-free formula G to $\exists x F$ in (21.12) is obvious, because each disjunct of the quantifier-free formula has a conjunct of the form F_x^t for some (extended) term t, even if it may be a square root expression or infinity or term involving infinitesimals. Whenever a formula of the form F_x^t is true, $\exists x F$ holds with that t as a witness, even when t is a square root expression, infinity, or infinitesimal.

The converse implication from $\exists x F$ to the quantifier-free formula G in (21.12) depends on showing that the quantifier-free formula G covers all possible representative cases and that the accompanying constraints on a, b, c, d are necessary so that they do not constrain solutions in unjustified ways.

One key insight is that it is enough to prove (21.12) for the case where all variables in F except x have concrete numeric real values, because the equivalence (21.12) is valid iff it is true in all states. So considering one concrete state at a time is enough. By a fundamental property of real arithmetic called *o-minimality*, the set

$$\mathscr{S}(F) = \{\omega(x) \in \mathbb{R} \,:\, \omega \in [\![F]\!]\}$$

of all real values for x that satisfy F forms a finite union of (pairwise disjoint) intervals, because the polynomials in F only change signs at their roots. There are only finitely many roots, now that the polynomials have become univariate, i.e., with the only variable x, since all free variables are evaluated to concrete real numbers in ω. Without loss of generality (by merging overlapping or adjacent intervals), all those intervals are assumed to be maximal, i.e., no bigger interval would satisfy F. So F actually changes its truth-value at most at the lower and upper endpoints of these intervals (unless the interval is unbounded). *Polynomials only change signs at their roots!*

The endpoints of these intervals are of the form $-c/b, (-b+\sqrt{d})/(2a), (-b-\sqrt{d})/(2a)$ or $\infty, -\infty$ for any of the polynomials ax^2+bx+c in F, because all polynomials in F are at most quadratic and all roots of those polynomials are of one of the above forms. In particular, if $-c/b$ is an endpoint of an interval of $\mathscr{S}(F)$

for a polynomial ax^2+bx+c in F, then $a=0, b\neq 0$, because that is the only case where $-c/b$ satisfies F, which has only at most quadratic polynomials. Likewise, if $(-b+\sqrt{d})/(2a)$ and $(-b-\sqrt{d})/(2a)$ are endpoints of intervals of $\mathscr{S}(F)$ for a polynomial ax^2+bx+c in F, then both imply that $a\neq 0$ and discriminant $d\geq 0$, otherwise there is no such solution in the reals. Consequently, all the side conditions for the roots in the quantifier-free formula G are necessary.

Now consider one interval $I\subseteq\mathscr{S}(F)$ (if there is none, $\exists x F$ is *false* and so will G be). If I has no lower bound in $\mathbb{R}$, then $F_{\mathrm{x}}^{-\infty}$ is true by construction (by Lemma 21.1, the virtual substitution $F_{\mathrm{x}}^{-\infty}$ is equivalent to the literal substitution $F_x^{-\infty}$ in $\pm\infty$-extended real arithmetic). Otherwise, let $\alpha\in\mathbb{R}$ be the lower bound of I. If $\alpha\in I$ (i.e., I is closed at the lower bound), then α is of the form $-c/b, (-b+\sqrt{d})/(2a), (-b-\sqrt{d})/(2a)$ for some equation $(ax^2+bx+c=0)\in F$ or some weak inequality $(ax^2+bx+c\leq 0)\in F$ from F. Since the respective extra conditions on a,b,c,d hold, the quantifier-free formula G evaluates to true. If, otherwise, $\alpha\notin I$ (i.e., I is open at the lower bound α), then α is of the form $-c/b, (-b+\sqrt{d})/(2a), (-b-\sqrt{d})/(2a)$ for some disequation $(ax^2+bx+c\neq 0)\in F$ or some strict inequality $(ax^2+bx+c<0)\in F$. Hence, the interval I cannot be a single point. So, one of the infinitesimal increments $-c/b+\varepsilon, (-b+\sqrt{d})/(2a)+\varepsilon$, or $(-b-\sqrt{d})/(2a)+\varepsilon$ is in $I\subseteq\mathscr{S}(F)$, because infinitesimals are smaller than all positive real numbers, so smaller than the interval length. Since the respective conditions a,b,c,d hold, the quantifier-free formula G is again true. Hence, in either case, the quantifier-free formula is equivalent to $\exists x F$ in state ω. Since the state ω assigning concrete real numbers to all free variables of $\exists x F$ was arbitrary, the same equivalence holds for all states ω, which means that the quantifier-free formula G is equivalent to $\exists x F$. That is $G\leftrightarrow\exists x F$ is valid, i.e., $\models G\leftrightarrow\exists x F$. □

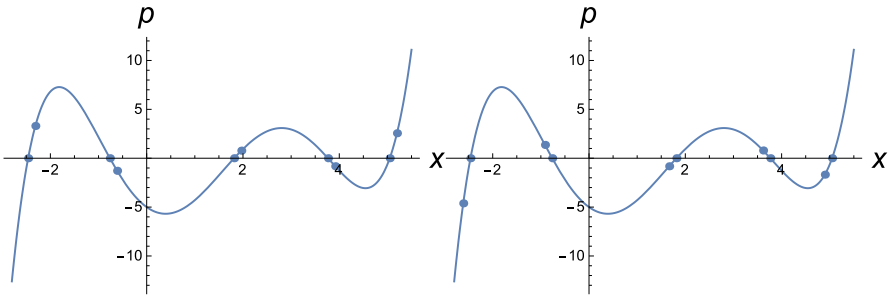


Fig. 21.3 Illustration of roots e and infinitesimal offsets $e+\varepsilon$ checked by virtual substitution along with $-\infty$ (**left**). Illustration of roots e and infinitesimal offsets $e-\varepsilon$ that could be checked along with $+\infty$ instead (**right**)

The order of the interval endpoints that the proof of Theorem 21.1 uses in addition to $-\infty$ is illustrated in Fig. 21.3(left). Observe that exactly one representative point is placed in each of the regions of interest, $-\infty$, each of the roots r, and just infinitesimally after the roots at $r+\varepsilon$. Alternatively, Theorem 21.1 could be

rephrased to work with ∞, at each root r, and always before the roots at $r-\varepsilon$; see Fig. 21.3(right) and Exercise 21.4. The illustrations in Fig. 21.3 show the ordering situation for a higher-degree polynomial p even if Theorem 21.1 only makes use of the argument for $p = ax^2+bx+c$ up to degree 2. Quantifier elimination procedures for higher degrees are still based on this fundamental principle, but require more subtle algebraic computations. The source of the trouble is Abel-Ruffini's impossibility theorem that there are, generally, no algebraic solutions to polynomial equations of degree ≥ 5. That is, the fact that we can characterize the roots of polynomials with roots was specific to degree ≤ 4 even when admitting nested roots.

Finally note that it is quite possible that the considered polynomial p does not single out the appropriate root e or off-root $e+\varepsilon$ that satisfies F to witness $\exists x F$. Then none of the points illustrated in Fig. 21.3 will satisfy F, because only a point other than $e+\varepsilon$ in the open interval between two roots will work.

Note 85 (No rejection without mention) The key argument underlying all quantifier elimination procedures in some way or another is that all parts of F that are not satisfied for any of the points in Fig. 21.3 that p brings about would have to mention another polynomial q with different roots $\tilde{e}$ and different off-roots $\tilde{e}+\varepsilon$ that will then enter the big disjunction in Theorem 21.1.

Example 21.3. The example of nonnegative roots of quadratic polynomials from Example 20.6 in Chap. 20 used Theorem 20.2 to construct and justify the quantifier elimination equivalence

$$\begin{aligned}&\mathrm{QE}(\exists x\,(ax^2+bx+c=0\wedge x\geq 0))\\ &\equiv b^2-4ac\geq 0\wedge(ba\leq 0\wedge ac\geq 0\vee a\geq 0\wedge ac\leq 0\vee a\leq 0\wedge ac\leq 0)\end{aligned}$$

under the assumption $a\neq 0$. Specializing to a case similar to Fig. 21.2 gives

$$\begin{aligned}\mathrm{QE}(\exists x\,(x^2-x+c=0\wedge x\geq 0))\equiv(-1)^2-4c\geq 0\wedge(c\geq 0\vee c\leq 0)\equiv 1-4c\geq 0\\ \equiv c\leq\frac{1}{4}\end{aligned}$$

By Theorem 21.1, the same square root expression substitution as in Example 20.6 in Chap. 20 will happen for the atomic formula $x^2-x+c\leq 0$ except that the case of $-\infty$ will be added as well as the root 0 that results from considering the linear atomic formula $-x\geq 0$:

$$\begin{aligned}&\mathrm{QE}(\exists x\,(x^2-x+c\leq 0\wedge x\geq 0))\equiv\\ &\underbrace{(x^2-x+c\leq 0\wedge\ldots)^{-\infty}_{\bar{x}}}_{false}\vee 1-4c\geq 0\vee\underbrace{(x^2-x+c\leq 0\wedge x\geq 0)^{0}_{\bar{x}}}_{c\leq 0\wedge 0\geq 0}\equiv 1-4c\geq 0\end{aligned}$$

Note that the additional disjunction $c\leq 0$ coming from the root 0 of $-x$ is in this case subsumed by the previous disjunct $1-4c\geq 0$. Hence, adding the roots of $-x$ did not modify the answer in this case. When adding a third conjunct $-x+2=0$,

this handling of all roots becomes critical:

$$\mathrm{QE}(\exists x\,(x^2-x+c\leq 0\wedge x\geq 0\wedge -x+2=0))$$

Since the first two polynomials x^2-x+c and $-x$ are still the same, the same virtual substitutions will happen as before. Except that they now fail on the new conjunct $-x+2=0$, because the root 0 of the polynomial $-x$ from the second conjunct does not satisfy $-x+2=0$ and because the virtual substitution of the roots $(-1\pm\sqrt{1-4c})/2$ of the first polynomial x^2-x+c fails:

$$(-x+2{=}0)_{\tilde{x}}^{(-1\pm\sqrt{1-4c})/2} \equiv ((1+\mp 1\sqrt{1-4c})/2{+}2{=}0) \equiv ((3+\mp 1\sqrt{1-4c})/2{=}0)$$
$$\equiv \mp 3\leq 0\wedge 3^2-(\mp 1)^2(1-4c)=0 \equiv -3\leq 0\wedge 3^2-(-1)^2(1-4c)=0 \equiv 8-4c=0$$

The latter is only possible for $c=2$, which is ruled out by the discriminant condition $1-4c\geq 0$ that precedes it. And, indeed, neither the roots of the quadratic polynomial illustrated in Fig. 21.2 nor the roots of $-x$ nor $-\infty$ are the right points to consider to satisfy the last conjunct. Of course, the last conjunct expresses that constraint by saying $-x+2=0$ quite explicitly. Never mind that this is an equation for now. Either way, the atomic formula clearly reveals that $-x+2$ is the polynomial that it cares about. So its roots might be of interest and will, indeed, by considered in the big disjunction of Theorem 21.1 as well. Since $-x+2$ is a visibly linear polynomial, its solution is $x=-2/-1=2$ which is even kind enough to be a standard real number so that literal substitution is sufficient and no virtual substitution is needed. Consequently, the substitution of this root $x=2$ of the last conjunct into the full formula quickly yields

$$(x^2-x+c\leq 0\wedge x\geq 0\wedge -x+2=0)_x^2 \equiv 2^2-2+c\leq 0\wedge 2\geq 0\wedge 0=0 \equiv 2+c\leq 0$$

This provides an answer that the quadratic polynomial x^2-x+c itself could not foresee because it depends on the polynomial $-x+2$ to even take this root into consideration. By Theorem 21.1, the overall result of quantifier elimination, thus, is the combination of the cases considered separately above:

$$\begin{aligned}
&\mathrm{QE}(\exists x\,(x^2-x+c\leq 0\wedge x\geq 0\wedge -x+2=0))\\
&\equiv \underbrace{(x^2-x+c\leq 0\wedge\ldots)_{\tilde{x}}^{-\infty}}_{\mathit{false}}\\
&\vee 1-4c\geq 0\wedge \underbrace{(\cdots\wedge -x+2=0)_{\tilde{x}}^{(-1\pm\sqrt{1-4c})/2}}_{8-4c=0}\\
&\vee -1\neq 0\wedge \underbrace{(x^2-x+c\leq 0\wedge x\geq 0)_x^0}_{c\leq 0\wedge 0\geq 0}\wedge \underbrace{(-x+2=0)_x^0}_{2=0}\\
&\vee -1\neq 0\wedge \underbrace{(x^2-x+c\leq 0\wedge x\geq 0\wedge -x+2=0)_x^2}_{2+c\leq 0} \qquad \equiv 2+c\leq 0\equiv c\leq -2
\end{aligned}$$

In this particular case, observe that Theorem 20.2 using $-x+2=0$ as the key formula would have been most efficient, because that would have gotten the answer right away without fruitless disjunctions. This illustrates that it pays off to pay attention with real arithmetic and always choose the computationally most parsimonious approach. But the example also illustrates that the same computation would happen if the third conjunct had been $-x+2\leq 0$, in which case Theorem 20.2 would not have helped.

21.6 Optimizations

Optimizations are possible for virtual substitutions [13] if there is only one quadratic occurrence of x, and that occurrence is not in an equation. If that occurrence is in an equation, Theorem 20.2 already showed what to do. If there is only one occurrence of a quadratic inequality, the following variation of Theorem 21.1 works, which uses exclusively linear fractions.

Note 86 ([13]) Let $\left(Ax^2+Bx+C\left\{\begin{smallmatrix}\leq\\<\\\neq\end{smallmatrix}\right\}0\right)\in F$ be the only quadratic occurrence of x. In that case, $\exists x F$ is equivalent over $\mathbb{R}$ to the following quantifier-free formula:

$$\begin{aligned}
&A=0\wedge B\neq 0\wedge F_{\bar{x}}^{-C/B}\vee A\neq 0\wedge F_{\bar{x}}^{-B/(2A)}\\
&\vee F_{\bar{x}}^{-\infty}\vee F_{\bar{x}}^{\infty}\\
&\vee \bigvee_{\left(0x^2+bx+c\left\{\begin{smallmatrix}=\\\leq\end{smallmatrix}\right\}0\right)\in F}\left(b\neq 0\wedge F_{\bar{x}}^{-c/b}\right)\\
&\vee \bigvee_{\left(0x^2+bx+c\left\{\begin{smallmatrix}\neq\\<\end{smallmatrix}\right\}0\right)\in F}\left(b\neq 0\wedge (F_{\bar{x}}^{-c/b+\varepsilon}\vee F_{\bar{x}}^{-c/b-\varepsilon})\right)
\end{aligned}$$

The clou in this case is that the extremal values of Ax^2+Bx+C are at the roots of the derivative

$$(Ax^2+Bx+C)'=2AX+B\stackrel{!}{=}0,\ \text{i.e.,}\ x=-\frac{B}{2A}$$

Since the only quadratic occurrence in Note 86 is not an equation, this extremal value is the only point of the quadratic polynomial that matters. In this case, $F_{\bar{x}}^{-B/(2A)}$ will substitute $-B/(2A)$ for x in the only quadratic polynomial as follows:

$$\left(Ax^2+Bx+C\left\{\begin{smallmatrix}\leq\\<\\\neq\end{smallmatrix}\right\}0\right)_{\bar{x}}^{-B/(2A)}\equiv\left(A\frac{(-B)^2}{4A^2}+\frac{-B^2}{2A}+C\left\{\begin{smallmatrix}\leq\\<\\\neq\end{smallmatrix}\right\}0\right)\equiv\left(\frac{-B^2}{4A}+C\left\{\begin{smallmatrix}\leq\\<\\\neq\end{smallmatrix}\right\}0\right)$$

The formula resulting from Note 86 might be bigger than that of Theorem 21.1 but *it does not increase the polynomial degree*, which can be crucial for nested quantifiers.

Further optimizations are possible if some signs of a, b are known, because several cases in the quantifier-free expansion then become impossible and can be simplified to *true* or *false* immediately. This helps simplify the formula in Theorem 21.1, because one of the cases $a = 0$ versus $a \neq 0$ might drop. But it also reduces the number of disjuncts in $F_{\tilde{x}}^{-\infty}$, see Example 21.1, and in the virtual substitutions of square roots (Chap. 20) and of infinitesimals (Sect. 21.4), which can lead to significant simplifications.

Theorem 21.1 also applies to polynomials of higher degrees in x if they factor to polynomials of at most quadratic degree in x [13]. Degree reduction is also possible by renaming based on the greatest common divisor of all powers of x that occur in F. If a quantified variable x occurs only with exponents that are multiples of an odd number d then virtual substitution can use $\exists x F(x^d) \equiv \exists y F(y)$. If x only occurs with degrees that are multiples of even number d then $\exists x F(x^d) \equiv \exists y (y \geq 0 \wedge F(y))$. It helps reduce the number of cases in Theorem 21.1 that infinitesimals $+\varepsilon$ are only needed if x occurs in strict inequalities in F. The cases $F_{\tilde{x}}^{(-b+\pm\sqrt{d})/(2a)}$ are only needed if x occurs in equations or weak inequalities.

21.7 Summary

Virtual substitution is one technique for eliminating quantifiers in real arithmetic. It works for linear and quadratic constraints and can be extended to some cubic cases [12]. Virtual substitution can be applied repeatedly from inside out to eliminate quantifiers. In each case, however, virtual substitution requires the eliminated variable to occur with small enough degree only. Even if that was the case initially, it may no longer be the case after eliminating the innermost quantifier, because the degrees of the formula resulting from virtual substitution may increase. In that case, degree optimizations and simplifications may sometimes work. If not, then other quantifier elimination techniques need to be used, which are based on semialgebraic geometry or model theory. Virtual substitution alone always works for mixed quadratic-linear formulas, i.e., those in which all quantified variables occur linearly except for one variable that occurs quadratically. In practice, however, many other cases turn out to work well with virtual substitution.

By inspecting Theorem 21.1 and its optimizations, we also observe that it is interesting to look at only closed sets or only open sets, corresponding to formulas with only $\leq$ and $=$ or formulas with only $<$ and $\neq$ conditions, respectively, because half of the cases then drop out of the expansion in Theorem 21.1. Furthermore, if the formula $\exists x F$ only mentions strict inequalities $<$ and disequations $\neq$, then all virtual substitutions will involve infinitesimals or infinities. While both are conceptually more demanding than virtual substitutions with mere square root expressions, the advantage is that both infinitesimals and infinities rarely satisfy any equations (except when they are trivial because all coefficients are zero). In that case, most formulas simplify tremendously. That is an indication in the virtual substitution method of

a more general phenomenon: existential arithmetic with strict inequalities or, dually, validity of universal arithmetic with weak inequalities, is computationally easier.

21.8 Appendix: Semialgebraic Geometry

The geometric counterparts of polynomial equations or quantifier-free first-order formulas with polynomial equations are affine varieties. The geometric counterparts of first-order formulas of real arithmetic that may mention inequalities are called semialgebraic sets in real algebraic geometry [1, 2]. By quantifier elimination, the class of sets definable with quantifiers is the same as the class of sets definable without quantifiers. Hence, the formulas of first-order real arithmetic exactly define semialgebraic sets.

Definition 21.3 (Semialgebraic Set). $S \subseteq \mathbb{R}^n$ is an *semialgebraic set* iff it is defined by a finite intersection of polynomial equations and inequalities or any finite union of such sets:

$$S = \bigcup_{i=1}^{t} \bigcap_{j=1}^{s} \{x \in \mathbb{R}^n \ : \ p(x) \sim 0\} \quad \text{where } \sim \in \{=, \geq, >\}$$

The geometric counterpart of the quantifier elimination result is that semialgebraic sets are closed under projection (the other closure properties are obvious in logic), which is the Tarski-Seidenberg theorem [10, 11].

Theorem 21.2 (Tarski-Seidenberg). *Semialgebraic sets are closed under finite unions, finite intersections, complements, and projection to linear subspaces.*

The semialgebraic sets corresponding to a number of interesting systems of polynomial inequalities are illustrated in Fig. 21.4.

Exercises

21.1. Consider the first-order real-arithmetic formula

$$\exists x\,(ax^2 + bx + c \leq 0 \wedge F) \tag{21.13}$$

The virtual substitution of the roots of $ax^2 + bx + c = 0$ according to Sect. 20.4 as well as of $-\infty$ according to Sect. 21.3 leads to

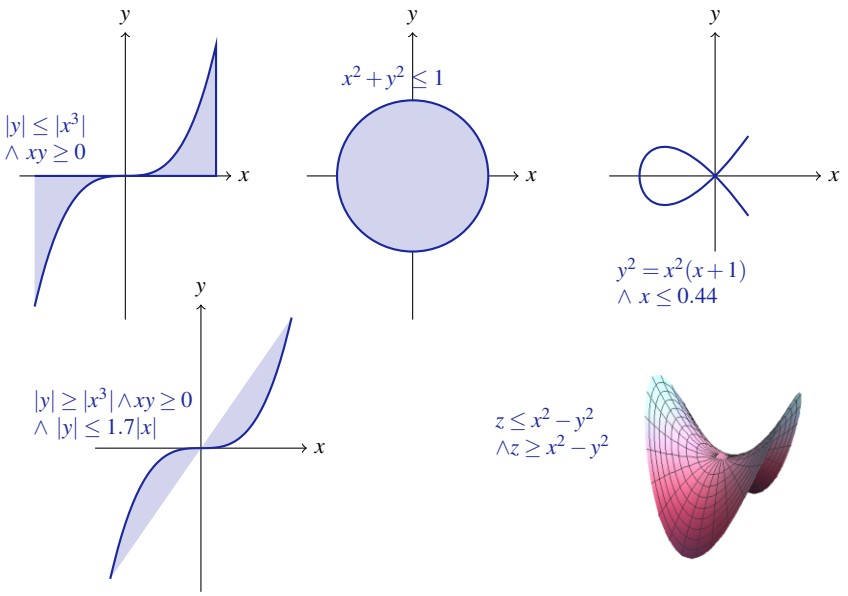


Fig. 21.4 Systems of polynomial inequalities describe semialgebraic sets

$$F_{\tilde{x}}^{-\infty} \vee a{=}0 \wedge b{\neq}0 \wedge F_{\tilde{x}}^{-c/b}$$
$$\vee\, a{\neq}0 \wedge b^2 - 4ac{\geq}0 \wedge \left(F_{\tilde{x}}^{(-b+\sqrt{b^2-4ac})/(2a)} \vee F_{\tilde{x}}^{(-b-\sqrt{b^2-4ac})/(2a)}\right)$$

But when F is $-ax^2 + bx + c < 0$, then none of those cases necessarily works. Does that mean the result of the virtual substitution is not equivalent to (21.13)? Where is the catch in this argument?

21.2. Perform quantifier elimination by virtual substitution to compute

$$\mathrm{QE}(\exists x\,(x^2 - x + c \leq 0 \wedge x \geq 0 \wedge -x + 2 \leq 0))$$

21.3. Consider the first-order real-arithmetic formula

$$\exists x\,(ax^2 + bx + c \leq 0 \wedge ax^2 + bx + c = 0)$$

Compare the results of using Theorem 20.1 and Theorem 21.1 on this formula. Which theorem is more efficient? What happens in the case of

$$\exists x\,(ax^2 + bx + c \leq 0 \wedge ax^2 + bx + c = 0 \wedge x \geq 0)$$

21.4 (Virtual substitution on the right). Develop and prove a virtual substitution formula for quadratic polynomials analogous to Theorem 21.1 that uses the points illustrated in Fig. 21.3(right) instead of Fig. 21.3(left).

21.5 (Infinitesimals in polynomials). Use the Taylor series

$$p(\zeta+\varepsilon) = \sum_{n=0}^{\infty} \frac{p^{(n)}(\zeta)}{n!}(\zeta+\varepsilon-\zeta)^n = \sum_{n=0}^{\infty} \frac{p^{(n)}(\zeta)}{n!}\varepsilon^n = \sum_{n=0}^{\deg(p)} \frac{p^{(n)}(\zeta)}{n!}\varepsilon^n$$

of univariate polynomial $p \in \mathbb{R}[x]$ around $\zeta \in \mathbb{R}$ evaluated at $\zeta+\varepsilon$ (since ε is small enough to be in the domain of convergence of the Taylor series) to show Note 83.

References

[1] Saugata Basu, Richard Pollack, and Marie-Françoise Roy. *Algorithms in Real Algebraic Geometry*. 2nd. Berlin: Springer, 2006. DOI: 10.1007/3-540-33099-2.

[2] Jacek Bochnak, Michel Coste, and Marie-Francoise Roy. *Real Algebraic Geometry*. Vol. 36. Ergeb. Math. Grenzgeb. Berlin: Springer, 1998. DOI: 10.1007/978-3-662-03718-8.

[3] Christopher W. Brown and James H. Davenport. The complexity of quantifier elimination and cylindrical algebraic decomposition. In: *ISSAC*. Ed. by Dongming Wang. New York: ACM, 2007, 54–60. DOI: 10.1145/1277548.1277557.

[4] Leonardo Mendonça de Moura and Grant Olney Passmore. Computation in real closed infinitesimal and transcendental extensions of the rationals. In: *Automated Deduction - CADE-24 - 24th International Conference on Automated Deduction, Lake Placid, NY, USA, June 9-14, 2013. Proceedings*. Ed. by Maria Paola Bonacina. Vol. 7898. LNCS. Berlin: Springer, 2013, 178–192. DOI: 10.1007/978-3-642-38574-2_12.

[5] Grant Olney Passmore. Combined Decision Procedures for Nonlinear Arithmetics, Real and Complex. PhD thesis. School of Informatics, University of Edinburgh, 2011.

[6] André Platzer. Differential dynamic logic for hybrid systems. *J. Autom. Reas.* **41**(2) (2008), 143–189. DOI: 10.1007/s10817-008-9103-8.

[7] André Platzer. *Logical Analysis of Hybrid Systems: Proving Theorems for Complex Dynamics*. Heidelberg: Springer, 2010. DOI: 10.1007/978-3-642-14509-4.

[8] André Platzer. Logics of dynamical systems. In: *LICS*. Los Alamitos: IEEE, 2012, 13–24. DOI: 10.1109/LICS.2012.13.

[9] André Platzer, Jan-David Quesel, and Philipp Rümmer. Real world verification. In: *CADE*. Ed. by Renate A. Schmidt. Vol. 5663. LNCS. Berlin: Springer, 2009, 485–501. DOI: 10.1007/978-3-642-02959-2_35.

[10] Abraham Seidenberg. A new decision method for elementary algebra. *Annals of Mathematics* **60**(2) (1954), 365–374. DOI: 10.2307/1969640.

[11] Alfred Tarski. *A Decision Method for Elementary Algebra and Geometry*. 2nd. Berkeley: University of California Press, 1951.

[12] Volker Weispfenning. Quantifier elimination for real algebra — the cubic case. In: *ISSAC*. New York: ACM, 1994, 258–263.

[13] Volker Weispfenning. Quantifier elimination for real algebra — the quadratic case and beyond. *Appl. Algebra Eng. Commun. Comput.* **8**(2) (1997), 85–101. DOI: 10.1007/s002000050055.