Software Theory Change for resilient near-complete specifications

Giuseppe Primiero\textsuperscript{a,*}, Franco Raimondi\textsuperscript{a}

\textsuperscript{a}Department of Computer Science, Middlesex University, London, UK

Abstract

Software evolution and its laws are essential for antifragile system design and development. In this paper we model early-stage \textit{perfective} and \textit{corrective} changes to software system architecture in terms of logical operations of expansion and safe contraction on a theory. As a result, we formulate an inference-based notion of property specification resilience for computational systems, intended as resistance to change. The individuated resilient core of a software system is used to characterize adaptability properties.

Keywords: Software Evolution; Theory Change; Property and System Resilience; System Adaptability.

1. Introduction

Software change is a critical step in the life-cycle of computational systems. Modifying or re-defining system specification properties is required by increasing architectural complexity or to improve software quality. In either case, “software maintenance has been regarded as the most expensive phase of the software cycle”\textsuperscript{35} p.32. A relevant amount of research has already been dedicated to the understanding, planning and execution of software evolution, in particular for requirements evolution, see e.g.\textsuperscript{16}. Typically, this occurs as part of the \textit{late} life-cycle of the system, dictated by architectural degeneration (violation or deviation of the architecture, increasing with changes being made to the original, see e.g.\textsuperscript{15,27}) and flexibility requirements (the system property that defines the extent to which the system allows for unplanned modifications\textsuperscript{32}). In this context, resilience as functionality preservation under changes is of paramount importance. The laws of software evolution for computational systems linked to the real environment\textsuperscript{24,25} express the importance of an appropriate understanding of software change. Change classification schemes, assessing the impact and risk associated with software evolution, present challenges\textsuperscript{28} which include integration in the conventional software development process model. This, in turn, means that a model of software change at design and implementation stages is essential to assess and anticipate errors and to determine system’s reliability in view of threats to functionalities.

Late life-cycle \textit{misfunctions}, where the system produces negative side-effects absent in other systems of the same type, require \textit{corrective changes}, which include testing, with the exception of model-based testing. \textit{Early} life-cycle
change can be classified by *disfunctions*, where the system is less reliable or effective than one expects in performing its function, see\textsuperscript{17}. At early design and implementation stages, *perfective changes* result from new or changed requirements, see\textsuperscript{26,33}. We explicitly ignore here the other two classifications, namely *adaptive changes* and *preventative changes*. A violation of the model specification in the implementation leads to a revision of the examined system. Such process of software change can be modelled similarly to scientific theory change, to account for both perfective and corrective changes in the implementation, when the latter is diverging from the model specification. We model software change as operations on a model of a scientific theory, defined according to the formal operations of AGM belief change, see\textsuperscript{2}. We constrain the revision operation to finite bases, as we assume that requirements specification of any however large software should be accounted for in terms of a finite representation. This area at the intersection of software engineering and theory change has been only very little explored: the only approach explicitly based on AGM is to be found in\textsuperscript{36}, offering a framework to reason about requirements evolution in terms of belief change operations. In\textsuperscript{10}, belief revision used to deal with change propagation in model evolution. In\textsuperscript{30}, Booth’s\textsuperscript{8} negotiation-style for belief revision is used to model change from current to to-be system requirements, aiming at some form of compromise based on prioritization. AGM belief revision has been investigated for logic programming under answer set semantics in\textsuperscript{13,14}. While notoriously a number of methods in software engineering have focused on developing implementation from specifications\textsuperscript{34,1}, our analysis concentrates on the modelling of both perfective and corrective changes to design new specifications from early (incorrect) implementations. We formalize such changes in terms of expansion and safe contraction operations. Most importantly, we use such operations to define property resilience and generalise it to a definition of system adaptability. The latter connection, in particular, appears to be entirely missing in the literature. The rest of this paper is structured as follows. In section 2 we offer an informal analysis of software as a theory, using the distinction between model and implementation; in section 3 we offer the formal analysis of requirements evolution as revision on such a theory; in section 5 we link this analysis to property resilience for software systems.

2. Software Theory

A *model* of a computational system $S_m$ is a representation of the system’s intended behaviour: as such, it naturally offers a representation of its specification in terms of law-like expressions formulating requirements of valid input states, and the result of (program) actions in output states. Identified preconditions guarantee *safety* of program execution, i.e. they include sufficient states for some program of type $p$ to obtain a state that satisfies the postconditions. Identifying preconditions that (may) lead to execution failure of an instance of program of type $p$ allows to define *completeness*. A complete description of $S_m$ is akin to what in software testing research is known as an *oracle*, a “procedure that determines what the correct behaviour of a system should be for all input stimuli with which we wish to subject the system under test”\textsuperscript{20}. A complete model $S_m$ of a correctly functioning computational system can be taken to mimic the notion of *right theory* for a software system. In this sense, $S_m$ is understood as the perfect description of design-for-test principles, which can be accompanied by a (possibly complete) formal specification of the intended behaviour. *Near-complete* specifications refer to approximations to complete specifications through the use of possible errors classification in weakened postconditions and an algebra on the possible program states, see\textsuperscript{23}. A similar sense of near-completeness is offered by approaches to requirements engineering where the relevant specification description can be marked up with refinements given by possible, admissible or acceptable malfunctioning. Where completeness is not achieved, a partial or near-complete oracle is a description that offers postconditions for some given inputs, and could offer alternative means for other inputs (metamorphic testing and derived information from example executions).

An *implementation* of $S_m$ is an actual realization of the model in some programming language. We assume that such realization is for all purposes and intentions faithful to the spirit of $S_m$. We assume moreover the possibility of abstracting (possibly automatically) from said implementation a set of law-like expressions of the same syntax as those forming $S_m$. We currently abstract from the concrete syntax of such expressions. Let us call this new set of specifications $S_i$. We assume, finally, that $S_i$ is in fact comparable to $S_m$, in the sense of containing at least some of the specification properties expressed by the latter. The comparison of requirements implemented in $S_i$ with $S_m$ can be performed in terms of log instances expressing observed truths of domain literals or task occurrences, where successful execution of the latter ones in appropriate order leads to goal satisfaction. A working implementation,
Faulty execution as expressed by divergence from expected or intended postconditions can be seen as a figure of drifting from fidelity, see\textsuperscript{12}. Reasons for labelling an instance of a program as malfunctioning with respect to its model can be reduced to either the requirement of stronger preconditions for running $p_i$ as intended by $S_m$, or the obtaining of different postconditions. Software systems' resistance to change in the environment as graceful degradation is a research topic of its own\textsuperscript{6,7}, including reference to hardware and material execution conditions, and we abstract from it in the present contribution to model specifically the relation of change in specifications as induced from unexpected postconditions. In the case of diverging descriptions of $S_i$ from $S_m$, perfective or corrective changes might be applied to produce a new model $S'_m$ that either accommodates the divergences, or eliminates some of the current (undesired) properties of the model $S_m$. This process, known as \textit{requirements evolution}, formally corresponds to a mapping $S_i \mapsto S'_m$. Informally, it can be associated with the restructuring of required properties in a model adapted to the current implementation conditions, a typically antifragile process. Formally, we require a non-monotonic logic which, in the case of perfective changes allows to add desired property-specifications and in the case of corrective changes allows to remove undesired ones. Comparing $S_i$ to $S_m$, it is possible to identify which elements of the former should be changed to have a better approximation to the latter, i.e. to reduce drifting and restore properties of the originally intended model in the implementation (possibly with a higher level of resistance). The formal operations described in the next section allow to establish which elements from $S_m$ are safe, when contraction of $S_i$ is performed.

3. Requirements Evolution by Theory Change

Consider a finite set of formulae $S_m = \{\phi_1, \ldots, \phi_n\}$, where each $\phi_i$ expresses a specific behaviour that the intended software system $S_m$ should display. The intended meaning of a formula $\phi_i$ is an instance of the relation over the state space represented by program actions defining $S_m$. We refer to $S_m$ as a theory of the model, i.e. its closure under logical implication $S_m = Cn(S_m) := \{\phi_i \mid S_m \models \phi_i\}$. We say that $S_m$ is consistent if $S_m \models \lnot(\phi_i \land \lnot\phi_i)$. We currently abstract away from the definition of the consequence relation $\models$ for $S_m$, which can be thought of as any classical consequence relation without completeness (which appears too strong in the context of real software specifications) and compactness (which is trivial in the present finite setting):

1. $S_m \models \top$
2. $S_m \models (\phi_i \rightarrow \phi_j)$ and $S_m \models \phi_i$, implies $S_m \models \phi_j$
3. $S_m \models \phi_i$ implies $S_m \not\models \lnot\phi_i$

$\models$ intuitively reflects property expressiveness: a formula $\phi_i \models \phi_j$ says that a property specification $\phi_i$ holding for a system $S_m$ \textit{induces} property specification $\phi_j$ in the corresponding theory $S_m$.

Consider now a new language $S_i$ for the model abstracted from an implementation of $S_m$ and so that for some $\phi_i$, either a theory $S_i = Cn(S_i)$ is such that $S_i \not\models \phi_i$ and $S_m \models \phi_i$; or $S_i \models \phi_i$ and $S_m \not\models \phi_i$. Formal operations can be defined on $S_i$ so that either the current input in the \textit{implementation} becomes valid, or the specification that makes our experimental result wrong is removed. In the former case, $S_i$ is expanded as to include $\phi_i$, hence one handles a form of requirement incompleteness: we indicate the result of this change as \textit{expansion}. This formal operation reflects the implementation of a new requirement and hence qualifies as a \textit{perfective change}. In the second case, $S_i$ is contracted as to remove $\phi_i$ (under a complete theory, which we do not assume, this induces satisfiability of $\lnot\phi_i$): the divergence between the implementation and the oracle reflects here a form of inconsistency, while at each stage of the implementation consistency is preserved; we indicate the result of this revision as \textit{contraction}. This formal operation
reflects the removal of an undesired requirement (error fixing) and hence qualifies as a corrective change. In both cases, a new model $S'_m$ is obtained, from which a new implementation can be formulated.

3.1. Expansion

The process of designing a piece of software can be seen as moving from an empty set of requirements (the trivial system specification, i.e. one that implements no operations) to one that includes some property specifications. This process is akin to an expansion of the software model abstracted from the trivial implementation $S_i = \emptyset$ with respect to a new specification requirement $\phi_i$, denoted as $S'_m = (S_i)^+_\phi_i$ and defined by the logical closure of $S_i$ and $\phi_i$. Theory creation has then a starting point $S_i \not\models \phi_i$, for any $\phi_i$. Any expansion operation after the first one should preserve consistency in $S_i$. Otherwise, each expansion by $\phi_i$ needs to be accompanied by the implicit elimination of the contradictory $\neg \phi_i$ from the list of feasible property descriptions according to $S_i$. Hence, each expansion requires a minimal set of contraction operations.

3.2. Contraction

We now consider contracting $S_i$ in view of a specification requirement $\phi_i$, inducing the removal of the minimal set of specifications implying $\phi_i$. We denote this by $(S_i)^{-}_{\phi_i} = \text{Cnt}(S_i, \phi_i)$, where the latter indicates the result of $S_i$ once $\phi_i$ has been removed. The contraction operator is then a mapping function:

$$S_i \mapsto S'_m := ((S_i, \phi_i)) \mapsto (S_i)^{-}_{\phi_i}$$

from the set of theories of the current $S_i$ to a new model whose implementations have languages that do not imply $\phi_i$. In the context of software engineering, a contraction operation should aim at removing the least expressive properties to induce a minimal loss of functionalities. We reflect this formally by an ordering $\leq$ on properties, similarly to what is done with epistemic entrenchment: $\phi_i \leq \phi_j$, in our context says that $\phi_j$ is at least as embedded as $\phi_i$ in the system in view of its functionalities. Hence, in a contraction process, one would remove first the latter in order to preserve as much as possible the operational properties of the system. $\leq$ satisfies the following postulates:

1. Transitivity: if $\phi_i \leq \phi_j$ and $\phi_j \leq \phi_k$, then $\phi_i \leq \phi_k$;
2. (Anti-)Dominance: if $\phi_i \not\models \phi_j$, then $\phi_j \leq \phi_i$;
3. Conjunctiveness: either $\phi_i \leq \phi_i \land \phi_j$ or $\phi_j \leq \phi_i \lor \phi_j$;
4. Minimality: if $S_i$ is consistent, then $\phi_i \not\in S_i$ iff $\phi_i \leq \phi_j$ for all $\phi_j$;
5. Maximality: if $\phi_j \not\in S_i$ for all $\phi_j$, then $\phi_i \in \text{Cnt}(\emptyset)$.

Standard Gärdenfors postulates are modified by (Anti-)Dominance, inverting the usual relation with $\models$. Among the different (although in some ways related, see\(^4\)) contraction functions, safe contraction is a natural candidate for the contraction on a finite set of property specifications under this ordering preserving system functionalities:

**Definition 3.1 (Safe Contraction,\(^3\)).** Given a theory $S_i$ of an implementation of $S_m$, ordered by a transitive non-circular relation $<$ (or hierarchy); let $\phi_i \in S_i$ express a property we wish to eliminate from the corresponding model $S_i$; then we say that a property $\phi_j \in S_m$ is safe with respect to $(S_i)^{-}_{\phi_i}$ modulo $<$, if and only if $\phi_j$ is not a minimal element under $<$ of any minimal $S'_m$ that verifies $\phi_i$.

Under such definition, $\phi_j$ is never the first property inducing $\phi_i$ in any sub-model $S'_m$ of $S_m$. Safe contraction satisfies the following rationality postulates:

1. Closure: $(S_i)^{-}_{\phi_i} \subseteq S_m \cap \text{Cnt}(S_i, \phi_i)$
2. Inclusion: $(S_i)^{-}_{\phi_i} \subseteq S_i$
3. Vacuity: $(\phi_i \not\in \text{Cnt}(S_i)) \rightarrow (S_i)^{-}_{\phi_i} = S_i$
4. Success: $(\phi_i \not\in \text{Cnt}(\emptyset)) \rightarrow \phi_i \not\in \text{Cnt}(S_i, \phi_i)$
5. Recovery: $(\phi_i \in \text{Cnt}(S_i)) \rightarrow S_i \subseteq (S_i)^{-}_{\phi_i}$
6. Extensionality: \((\phi_i \equiv \phi_j) \rightarrow (S_i)^{-}_{\phi_i} = (S_j)^{-}_{\phi_j}\)

Along the lines of the interpretation of \(\leq\) in terms of security and reliability in\(^3\), if the consequence relation \(\models\) for \(S_i\) is intended to describe specification expressiveness, then the more it can be logically inferred from a property, the more expressive that property is. In turn, our safe contraction module \(\leq\) makes more expressive properties safer, removing first those with the least inferential impact. This justifies our (Anti-)Dominance axiom; with Transitivity, the following counter-continuing conditions hold\(^3\):

**Proposition 3.1 (Counter-Continuing Down).** If \(\phi_i < \phi_j\) and \(\phi_k \models \phi_j\), then \(\phi_i < \phi_k\), for all \(\phi_i, \phi_j, \phi_k \in S_i\). In other words, if \(\phi_i\) is less safe for contraction than \(\phi_j\) (hence the former should be easier to remove than the latter) and \(\phi_k\) is more inferentially powerful than \(\phi_j\) (hence the former should be harder to remove than the latter), then \(\phi_i\) is also less safe for contraction than \(\phi_k\).

**Proposition 3.2 (Counter-Continuing Up).** If \(\phi_i \models \phi_j\) and \(\phi_i < \phi_k\), then \(\phi_j < \phi_k\), for all \(\phi_i, \phi_j, \phi_k \in S_i\). This says that if \(\phi_i\) is inferentially more powerful than \(\phi_j\) (hence the former should be harder to remove than the latter), and \(\phi_i\) is also strictly less safe from contraction than \(\phi_k\), then \(\phi_j\) is strictly more safe from contraction than \(\phi_k\).

### 3.3. Revision

Safe contraction leads to the understanding of revision of a software model \(S_i\) with respect to a new (possibly inconsistent) specification requirement \(\phi_i\) denoted as \(S_i^*\) as the accommodation of \(\phi_i\) involving as little change as possible. The revision operator is then a mapping function

\[
S_i \mapsto S_m := \{(S_i, \phi_i)\} \mapsto (S_i)^*_{\phi_i}
\]

from the set of theories of \(S_i\) to a new model \(S_m^*\) whose implementations have languages that imply \(\phi_i\). Counterparts to the above rationality postulates hold for revision. According to Levi identity, revision is equivalent to contraction followed by expansion: \(S_i^* = (S_i^\neg)_{\phi_i}^+\).

### 4. Example

It has been recently shown that most Mergesort algorithms are broken, including the Timsort hybrid algorithm\(^19\). We present here briefly the specification evolution from the broken implementation to the fixed specification, with remarks adapted to our analysis. The main loop of Timsort:

```java
do {
    int runLen = countRunAndMakeAscending(a, lo, hi, c);
    if (runLen < minRun)
    {
        int force = nRemaining <= minRun ? nRemaining : minRun;
        binarySort(a, lo, lo + force, lo + runLen, c);
    runLen = force;
    }
    ts.pushRun(lo, runLen);
    ts.mergeCollapse();
    lo += runLen;
    nRemaining = runLen;
} while (nRemaining != 0);
assert lo == hi;
assert ts.stackSize == 1;
```

can be represented as a theory \(S_m\), which among others satisfies a formula \(\phi_i\), which is an instance of the main loop with \(\text{stackSize}= 4\). A Java implementation \(S_i\) shows that the algorithm is broken with respect to such \(\phi_i\) after violation of the invariant, `ArrayIndexOutOfBoundsException` in `pushRun`, see\(^19\):
private void mergeCollapse() {
    while (stackSize > 1) {
        int n = stackSize - 2;
        if (n > 0 && runLen[n-1] <= runLen[n] + runLen[n+1]) {
            if (runLen[n - 1] < runLen[n + 1])
                n--;
            mergeAt(n);
        } else if (runLen[n] <= runLen[n + 1]) {
            mergeAt(n);
        } else {
            break; // Invariant is established
        }
    }
}

private void newMergeCollapse() {
    while (stackSize > 1) {
        int n = stackSize - 2;
        if (n > 0 && runLen[n-1] <= runLen[n] + runLen[n+1] ||
            n-1 > 0 && runLen[n-2] <= runLen[n] + runLen[n-1]) {
            if (runLen[n - 1] < runLen[n + 1])
                n--;
        } else if (n<0 || runLen[n] > runLen[n + 1]) {
            break; // Invariant is established
        }
        mergeAt(n);
    }
}

A new corrected model $S'_m$ is obtained by contraction of the merging step at a too low entry in the stack and by expansion to specify the right entry in the remaining stack where the merge happens. The new implementation $S'_i$ removes mergeAT(n) commands and $<=$ predicates and adds OR clauses and mergeAT(n) commands in the appropriate loops:

private void newMergeCollapse() {
    while (stackSize > 1) {
        int n = stackSize - 2;
        if (n > 0 && runLen[n-1] <= runLen[n] + runLen[n+1]) {
            if (runLen[n - 1] < runLen[n + 1])
                n--;
            mergeAt(n);
        } else if (runLen[n] <= runLen[n + 1]) {
            mergeAt(n);
        } else {
            break; // Invariant is established
        }
    }
}

Notice that merging of the last 3 elements of runLen is preserved, while the corresponding merging is not.

5. Inferential-based Resilience for System Adaptability

Resilience for a computational system reflects its (graded) ability to preserve a working implementation under varied specifications. The above analysis of software theory change allows us to provide a precise definition of resilience in the presence of failing components. In the literature on software change, this process corresponds to preservation of behavioural safety by specification approximation, see e.g. the taxonomy offered in9. Various attempts have been made to formalise perseverance of validity to change. The most common one encountered in this research area is that of system robustness. One (older) interpretation is given in terms of the inability of the system to distinguish between behaviours that are essentially the same, see31. More recently, the term resilience has been used to refer to the ability of a system to retain functional and non-functional identity with the ability to perceive environmental changes; to understand their implications and to plan and enact adjustments intended to improve the system-environment fit11. In5, a value-based notion of resilience is presented, according to which a prompt in a design process is resilient if it is a feature resisting all value-based relations of currently involved stakeholders and an Information System is called strongly resilient if, in view of possibly conflicting value-based future configurations, the system admits of new relations accommodating them. Resilience of software system specifications evaluated with respect to resistance to contraction operations in the relevant model induces a relation between inferential power of specifications and stronger system behaviour. Accordingly, their removal is more dangerous for system’s functionality. Then the following inferential-based definition of resilience can be formulated:

Definition 5.1 (Property Resilience). Consider property specifications $\phi_{i,j,k} \in S_m$ and a relevant implementation $S_i$. Then $\phi_i$ is more resilient than $\phi_j$ with respect to $S'_m = (S_i)^\phi_i$ if it is not a maximally vulnerable element of some minimal $S'_m \subseteq S_m$ implying $\phi_i$. 
Accordingly, one can generalize to system resilience:

**Definition 5.2 (System Resilience).** A software system specification $S_m$ is said resilient with respect to a property specification $\phi$ if the latter is not a maximally vulnerable element in any $S'_m$ preserving minimal functionalities of $S_m$.

System resilience as the resistance to change of property specifications (as in Definition 5.1) can be essential to determine system antifragility. Software antifragility has been characterized as self-healing (automatic run-time bug fixing) and adaptive fault-tolerance (tested e.g. by fault-injection in production)\(^2\). An inferential notion of resilience helps characterizing a certain degree of fault-tolerance; the latter is considered strictly intertwined with self-healing properties: while not all fault-tolerant systems are self-healing, one can argue that self-healing techniques are ultimately dependable computing techniques\(^2\). Our resilient core, intended as the persistence of service delivery\(^2\), allows to determine the adaptation required by changes in terms of valid and invalid properties of its contractions; and can anticipate results of its expansions.

The first property establishes that, given the non-resilient part of the system, it is possible to establish which properties will not be instantiated in any subsystem.

**Proposition 5.1 (Accountability).** For any maximally vulnerable $\phi_i$ of some $S'_m \subset S_m$, if $\phi_i \models \phi_j$, then $S'_m \not\models \phi_j$.

The second property establishes that, given the resilient core of the system, it is possible to establish which properties will be instantiated by any subsystem.

**Proposition 5.2 (Evolvability).** For any non-maximally vulnerable $\phi_i$ of some $S'_m \subset S_m$, if $\phi_i \models \phi_j$, then $S'_m \models \phi_j$.

The third property establishes that, given the resilient core of the system, it is possible to know which properties will be instantiated by any of its extensions.

**Proposition 5.3 (Prevision).** For any non-maximally vulnerable element $\phi_i$ of some $S'_m \subset S_m$, if $\phi_i \models \phi_j$ then there is $S''_m \subseteq S_m$ such that $S''_m \models \phi_j$.

6. Conclusions

In this paper we have considered software systems as theories, whose implementations show possibly diverging postconditions from the intended specification, and modelled changes following the AGM paradigm. A definition of dynamic resilience for such systems results from safe contraction operations that focus on preserving properties of the originally intended model, while accommodating changes and make it possible to anticipate required adaptation. In future research, we will consider the update operation from the AGM model to formulate adaptive changes as responses to variations in the environment.

Acknowledgements

We wish to thank Nikos Gorogiannis, Balbir Barn and Kelly Androutsopoulos for their input.

References
