



Matching in Quantitative Equational Theories

Georg Ehling, Temur Kutsia

► To cite this version:

Georg Ehling, Temur Kutsia. Matching in Quantitative Equational Theories. UNIF 2023 - 37th International Workshop on Unification, Veena Ravishankar; Christophe Ringeissen, Jul 2023, Rome, Italy. hal-04128216

HAL Id: hal-04128216

<https://inria.hal.science/hal-04128216>

Submitted on 14 Jun 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License

Matching in Quantitative Equational Theories

Georg Ehling and Temur Kutsia

RISC, Johannes Kepler Universität Linz, Hagenberg, Austria
{gehling, tkutsia}@risc.jku.at

Abstract

Quantitative equational reasoning, as introduced by Mardare, Panangaden, and Plotkin (2016), provides a framework that extends the equality predicate on terms to a notion of proximity. It allows to consider “indexed equations” $s \approx_\varepsilon t$ between terms with the idea of expressing that “ s is within ε from t ”.

In this work-in-progress paper, we discuss matching in the setting of quantitative equational reasoning. We outline a matching algorithm for a special case and suggest possible generalizations.

1 Introduction

Extending the equality predicate to a relation that expresses similarity or proximity is a task that has been addressed in various ways. Among the more prominent approaches, fuzzy reasoning equips an equation with a real number between 0 and 1 that expresses the degree to which it holds true. Equational problems like matching and unification in the fuzzy setting have been studied [e.g. Ses02; JR15; Pau22; KP19; PK21].

In [MPP16], the authors introduce quantitative equational reasoning as a framework that extends the equality predicate to a notion of proximity in the sense of metric spaces. They consider “indexed equations” $s \approx_\varepsilon t$ between terms with the idea of expressing that “ s is within a distance of ε from t ”, where ε is a nonnegative rational number. They provide semantic interpretations as well as a set of inference rules for quantitative equational logic and prove a quantitative analogue of Birkhoff’s theorem on the soundness and completeness of equational logic.

In this paper, we discuss matching in the setting of quantitative equational reasoning. We outline a matching algorithm for the special case where the theory is induced by a set of regular quantitative equations¹ with strictly positive index, present several examples and suggest possible generalizations.

2 Notation and Assumptions

We assume the reader’s familiarity with the basic notions of equational reasoning [e.g., see BN98]. The terminology and notation for quantitative equational reasoning used here is mostly guided by the one in [MPP16] and [Bac+20].

To begin with, fix a signature Ω and a countably infinite set of variables X . We denote by $T(\Omega, X)$ the set of terms over Ω and X , and by $\Sigma(X)$ the set of substitutions on X . Unless stated otherwise, we will use the letters a, b, c to denote constants, f, g, h for non-constant function symbols, l, r, s, t, u, v for terms and x, y, z for variables.

We will consider the quantitative equational theory \mathcal{U}_E axiomatized by a set of quantitative judgements E , as defined in [Bac+20, Section 10.4.1]. In this paper, we work under the running

¹An equation $s \approx_\varepsilon t$ is called regular if s and t have the same set of variables: $\text{var}(s) = \text{var}(t)$.

assumption that E is finite and *unconditional*, that is, that E contains exclusively judgements of the form $\emptyset \vdash s \approx_\varepsilon t$ for $s, t \in T(\Omega, X)$ and $\varepsilon \in \mathbb{Q}_{\geq 0}$. To keep the notation compact, we will write $s \approx_\varepsilon t \in E$ instead of $\emptyset \vdash s \approx_\varepsilon t \in E$. Moreover, we will use the notation $E \vdash s \approx_\varepsilon t$ to express that $\emptyset \vdash s \approx_\varepsilon t \in \mathcal{U}_E$.

We collect equations from E in two separate sets, depending on whether their index is zero or positive:

$$E_0 = \{s \approx_\varepsilon t \in E \mid \varepsilon = 0\}, \quad E_+ = \{s \approx_\varepsilon t \in E \mid \varepsilon > 0\}.$$

We sometimes refer to E_0 and E_+ as the *crisp* and *quantitative* part of E , respectively. For notational convenience, we moreover identify \approx and \approx_0 , i.e. we do not distinguish between E_0 and its non-quantitative counterpart $E'_0 = \{s \approx t \mid s \approx_0 t \in E_0\}$.

Analogous to the non-quantitative case, we say that a quantitative equation $s \approx_\varepsilon t$ is *regular* if s and t contain the same variables.

Throughout the paper, we will refer to some of the rules defining quantitative algebraic theories, which can be found in [Bac+20, Section 10.4.1].

Remark 2.1. A first simple, yet important observation is that we can assume all indices of quantitative equations in our setting to be integers (since multiplying all indices by a suitable rational to cancel denominators yields a theory that behaves analogous to the original one). Thus, induction on the index is available for future proofs.

3 Matching in the Quantitative Setting

3.1 Problem Formulation

Let Ω, X and E be as above and let $s, t \in T(\Omega, X)$ be terms, ε a nonnegative rational number. The *fixed-range* quantitative matching problem $s \lesssim_{E, \varepsilon}^? t$ (or simply $s \lesssim_\varepsilon^? t$, if E is clear from the context) consists in the following task:

($s \lesssim_{E, \varepsilon}^? t$): Find a substitution $\sigma \in \Sigma(X)$ such that $E \vdash s\sigma \approx_\varepsilon t$ (if such a σ exists).

A similar task, which will however not be studied in this paper, is presented by the *unrestricted* quantitative matching problem, where the distance is not fixed in advance:

($s \lesssim_{E, ?}^? t$): Find the least rational $\delta \geq 0$ such that there exists a substitution σ satisfying $E \vdash s\sigma \approx_\delta t$.

3.2 First Steps Towards a Solution

In order to solve the (fixed-range) matching problem, it would be sufficient to compute all elements in the “ball” $B_\varepsilon(t) := \{u \in T(\Omega, X) \mid E \vdash t \approx_\varepsilon u\}$ and syntactically match s to each of them. However, $B_\varepsilon(t)$ need not be finite for two reasons, as illustrated by the example below:

Example 3.1. (1) Take $E = \{f(x) \approx_1 g(x, y)\}$ and $t = f(a)$. Then $E \vdash f(a) \approx_1 g(a, y)$ by the (Subst) rule, and thus, every instance of $g(a, y)$ is in $B_1(f(a))$, again by (Subst). In particular, $B_1(t)$ is infinite.

(2) Take $E = \{x \approx_0 f(x)\}$, and let $t = a$ be a constant. Then by the (Triang) rule, we have $f^n(a) \in B_0(a)$ for every n , so $B_0(t)$ is infinite.

In the first example, the infinity of the ball is caused by the fact that E_+ allows to introduce new variables, which can be removed by imposing regularity of E_+ . In the second example, the problem lies in the nature of E_0 . We will see later (in Prop. 3.6) that all balls are finite if $E_0 = \emptyset$ and E_+ is regular. For the general case however, we should account for those problems by computing a compact representation of the ball, that is, one that is still finite but yet in some sense contains a representative of every relevant term. For the first type of problem, this means that the set we construct will only contain the non-ground generalization, but not its instances. For the second type of problem, it means that our set will only contain a representative of the term in the ball up to E_0 .

We define this compact representation of the ball as follows:

Definition 3.2. Let $E = E_0 \cup E_+$ be such that E_+ is finite with integer indices and E_0 has finitary unification type. Moreover, assume that a unification algorithm for E_0 is given. Given a term t and a rational $\varepsilon \geq 0$, we define the set $\mathcal{R}_\varepsilon(t)$ recursively by setting $\mathcal{R}_\varepsilon(t) = \{t\}$ if t is a variable and

$$\mathcal{R}_\varepsilon(t) = \{t\} \cup \bigcup_{\substack{\delta \in \mathbb{N}, \\ 0 < \delta \leq \varepsilon, \\ t = f(t_1, \dots, t_n), \\ s_i \in \mathcal{R}_\delta(t_i)}} \mathcal{R}_{\varepsilon - \delta}(f(s_1, \dots, s_n)) \cup \bigcup_{\substack{l \approx_\delta r \in \bar{E}_+, \\ \delta \leq \varepsilon, \\ \sigma \in \text{mcu}_{E_0}(l, t)}} \mathcal{R}_{\varepsilon - \delta}(r\sigma) \quad (1)$$

otherwise, where “ $l \approx_\delta r \in \bar{E}_+$ ” means that either $l \approx_\delta r$ or $r \approx_\delta l$ is a fresh variant of an equation in E_+ , and where $\text{mcu}_{E_0}(l, t)$ stands for a minimal complete set of E_0 -unifiers of l and t computed by the given algorithm.

Remark 3.3. (1) If t is ground, then $\mathcal{R}_\varepsilon(t)$ is a finite set. It is defined uniquely up to renaming of variables. Fixing a unification algorithm for E_0 serves the sole purpose of guaranteeing this uniqueness.

(2) As announced earlier, the idea of the above definition is to represent all terms that are ε -close to t . The first set in (1), the singleton $\{t\}$, guarantees that t itself is in the representation; this corresponds to the reflexivity rule (Refl) of quantitative equational theories. The second set makes sure that any term with the same head symbol as t whose arguments are close to those of t is close to t , reflecting the (NExp) rule. The union constituting the rest of the definition is inspired by term rewriting. Here, t is rewritten according to a rule from E_+ to some $r\sigma$, and we include elements that are $(\varepsilon - \delta)$ -close to $r\sigma$, since a distance of δ has been consumed by the rewriting process. We will illustrate in Example 3.5 why it makes sense to consider a unifier σ here instead of a matcher as in classical term rewriting. This part reflects the (Symm) and (Triang) rules.

Example 3.4. Consider $E = \{f(x, y) \approx_0 f(y, x), f(x, y) \approx_1 g(x, y)\}$ and the matching problem $f(g(b, z), z) \lesssim_1 f(f(a, b), a)$.

Let t be $f(f(a, b), a)$ for short. We show the computation of $\mathcal{R}_1(t)$ in full detail. Note that E_0 -unification between t and the right hand side of $f(x, y) \approx_1 g(x, y)$ fails, and for the left hand side, the minimal complete set of E_0 -unifiers depends on the choice of the E_0 -unification algorithm that was fixed beforehand. We choose here to compute $\mathcal{R}_1(t)$ with respect to the unification algorithm that gives $\sigma_1 = \{x \mapsto a, y \mapsto f(a, b)\}$ and $\sigma_2 = \{x \mapsto f(a, b), y \mapsto a\}$ as the minimal complete set of unifiers of t and $f(x, y)$. Then, we have

$$\mathcal{R}_1(t) = \{t\} \cup \bigcup_{\substack{u \in \mathcal{R}_1(f(a, b)), \\ v \in \mathcal{R}_1(a)}} \mathcal{R}_0(f(u, v)) \cup \bigcup_{i \in \{1, 2\}} \mathcal{R}_0(g(x, y)\sigma_i).$$

Since $\mathcal{R}_0(s) = \{s\}$ holds for any term s , this simplifies to

$$\mathcal{R}_1(t) = \{t\} \cup \{f(u, v) \mid u \in \mathcal{R}_1(f(a, b)), v \in \mathcal{R}_1(a)\} \cup \{g(a, f(a, b)), g(f(a, b), a)\}.$$

We now need to compute the sets $\mathcal{R}_1(f(a, b))$ and $\mathcal{R}_1(a)$. Since a cannot be unified modulo E_0 with a side of $f(x, y) \approx_1 g(x, y)$, we obtain $\mathcal{R}_1(a) = \{a\}$. As for the computation of $\mathcal{R}_1(f(a, b))$, note that the minimal complete set of unifiers of $f(a, b)$ and $f(x, y)$ is given by $\{x \mapsto a, y \mapsto b\}$, $\{x \mapsto b, y \mapsto a\}$, and we obtain

$$\begin{aligned} \mathcal{R}_1(f(a, b)) &= \{f(a, b)\} \cup \bigcup_{\substack{u \in \mathcal{R}_1(a) \\ v \in \mathcal{R}_1(b)}} \mathcal{R}_0(f(u, v)) \cup \mathcal{R}_0(g(a, b)) \cup \mathcal{R}_0(g(b, a)) \\ &= \{f(a, b), g(a, b), g(b, a)\}. \end{aligned}$$

Hence, the computation gives

$$\mathcal{R}_1(t) = \{t, f(g(a, b), a), f(g(b, a), a), g(a, f(a, b)), g(f(a, b), a)\}.$$

The solution to the matching problem $f(g(b, z), z) \lesssim_1 f(f(a, b), a)$ is now found as the matcher $\tau = \{z \mapsto a\}$ of $f(g(b, z), z)$ to $f(g(b, a), a) \in \mathcal{R}_1(t)$.

Example 3.5. Consider $E = \{f(x, y) \approx_1 g(x), f(x, a) \approx_1 h(x)\}$ and the matching problem $h(x) \lesssim_2 g(b)$. We have $\mathcal{R}_2(g(b)) = \{g(b)\} \cup \mathcal{R}_1(f(b, y)) = \{g(b), f(b, y), h(b)\}$, and the solution $\{x \mapsto b\}$ to the matching problem can be found as the syntactic matcher of $h(x)$ to $h(b)$. Note that the presence of $h(b)$ in $\mathcal{R}_2(g(b))$ hinges on the fact that unification was used in the definition of $\mathcal{R}_\varepsilon(t)$: if matching had been used instead, $h(b)$ would go missing because $f(x, a)$ cannot be matched to $f(b, y)$.

3.3 First Results

We now present some properties the compact representation $\mathcal{R}_\varepsilon(t)$ of the ball of radius ε around t enjoys. Let r, s, t be terms, ε, δ nonnegative rationals. The following can be shown by well-founded induction:

- (a) $\mathcal{R}_0(t) = \{t\}$
- (b) If $\delta \leq \varepsilon$, then $\mathcal{R}_\delta(t) \subseteq \mathcal{R}_\varepsilon(t)$.
- (c) Assume that $E_0 = \emptyset$, E_+ is regular and s, t are ground. Then $s \in \mathcal{R}_\varepsilon(t)$ if and only if $t \in \mathcal{R}_\varepsilon(s)$. Moreover, if $r \in \mathcal{R}_\delta(s)$ and $s \in \mathcal{R}_\varepsilon(t)$, then $r \in \mathcal{R}_{\delta+\varepsilon}(t)$.

Note that these properties reflect some of the rules defining quantitative equational theories [cf. [Bac+20](#), p. 347]. For instance, (c) reflects the (Symm) and (Triang) rules.

Both conditions in (c) are necessary: For instance, if $E_0 = \emptyset$ and $E_+ = \{f(x) \approx_1 g(x, y)\}$, then $f(a) \in \mathcal{R}_1(g(a, b))$, but $g(a, b) \notin \mathcal{R}_1(f(a))$. On the other hand, if $E_0 = \{f(x, y) \approx_0 f(y, x)\}$ and $E_+ = \{g(x) \approx_1 f(a, x)\}$, then $g(b) \in \mathcal{R}_1(f(b, a))$, but $f(b, a) \notin \mathcal{R}_1(g(b))$.

The goal of introducing $\mathcal{R}_\varepsilon(t)$ was to provide a compact representation of the ball $B_\varepsilon(t)$. More concretely, in the simplest scenario, this means the following:

Proposition 3.6. *Let $E = E_+$ be finite and regular, and let t be a ground term. Then $B_\varepsilon(t) = \mathcal{R}_\varepsilon(t)$.*

Algorithm 1: Quantitative matching for empty E_0 and regular E_+

Input : A finite set $E = E_+$ of quantitative equations with positive index, terms s, t with t ground.
Output: A complete set S of solutions to the matching problem $s \lesssim_{E, \varepsilon}^? t$
 $S \leftarrow \emptyset$;
compute $\mathcal{R}_\varepsilon(t)$;
for $u \in \mathcal{R}_\varepsilon(t)$ **do**
| $S \leftarrow S \cup \{\text{syntactic matchers of } s \text{ to } u\}$
end
return S

Proposition 3.6 gives us an algorithm to solve the matching problem $s \lesssim_\varepsilon t$ for regular E_+ and empty E_0 : Compute $\mathcal{R}_\varepsilon(t)$, and syntactically match s to each of its elements.

The soundness and completeness of Algorithm 1 are direct consequences of Proposition 3.6. Note that in a practical implementation, the computation could be sped up by interlacing the computation of the representation of the ball and syntactical matching, e.g. to avoid computing elements of $\mathcal{R}_\varepsilon(t)$ to which s cannot be matched.

3.4 Relaxing the assumptions

As seen in Example 3.1, Proposition 3.6 in general does not hold without the assumption of empty E_0 and regular E_+ . Let us investigate how Algorithm 1 could possibly be adapted in these cases. We examine first the case where E_0 is empty, but E_+ need not be regular. Recall that $\mathcal{R}_\varepsilon(t)$ in general does not contain the ground instances of non-ground terms that arise from non-regularity. Thus, in Algorithm 1, syntactic unification needs to be employed when comparing s to the elements of $\mathcal{R}_\varepsilon(t)$ rather than syntactic matching. This is illustrated by Example 3.7 below.

Example 3.7. Consider $E = \{f(x) \approx_1 g(x, y)\}$ and the matching problem $g(x, b) \lesssim_{E, 1}^? f(a)$. Then $\mathcal{R}_1(f(a)) = \{f(a), g(a, y)\}$.

Syntactic matching of $g(x, b)$ to elements of $\mathcal{R}_1(f(a))$ does not succeed. The solution $\sigma = \{x \mapsto a\}$ can be found via syntactic unification of $g(x, b)$ and $g(a, y)$.

Considering now the case where E_0 need not be empty, recall that $\mathcal{R}_\varepsilon(t)$ only records terms up to equality modulo E_0 . Thus, it should be expected that syntactic matching needs to be replaced by E_0 -unification in Algorithm 1:

Example 3.8. Consider $E = \{f(a, x) \approx_1 g(x, a), a \approx_0 b\}$ and the problem $f(b, y) \lesssim_1 g(c, b)$. Then $\mathcal{R}_1(g(c, b)) = \{g(c, b), f(a, c)\}$.

The solution $\sigma = \{y \mapsto c\}$ can be found as an E_0 -unifier of $f(b, y)$ and $f(a, c)$.

However, problems can occur in cases like the following:

Example 3.9. Take $E = \{f(a, x) \approx_0 g(x), a \approx_1 b\}$, and consider the matching problem $f(b, y) \lesssim_1 g(a)$. Since $E \vdash f(b, a) \approx_1 f(a, a) \approx_0 g(a)$, the substitution $\sigma = \{y \mapsto a\}$ is a solution. However, no element of $\mathcal{R}_1(g(a)) = \{g(a), g(b)\}$ is unifiable modulo E_0 with $f(b, y)$.

To find the matcher, one would need to consider also $\tilde{\mathcal{R}}_1(f(b, y)) = \{f(b, y), f(a, y)\}$, and note that σ is an E_0 -unifier of $f(a, y)$ and $g(a)$. Here, the notation $\tilde{\mathcal{R}}_1(f(b, y))$ should express

that it is not allowed to substitute for the variable y during the computation of the representation of the ball. This could be realized e.g. by first replacing y by a new free constant c_y , then computing $\mathcal{R}_1(f(b, c_y))$ and afterwards replacing every occurrence of c_y by y again.

4 Outlook

While Proposition 3.6 only provides an algorithmic solution to the quantitative matching problem in the very restricted case that $E_0 = \emptyset$ and E_+ is regular, we hope that similar ideas can be used to solve the matching problem in a more general setting, as outlined in Subsection 3.4. Establishing the corresponding results is an obvious next step.

Possible further generalizations could include removing the restriction to unconditional equations which we imposed in this paper as well as investigating the unrestricted matching problem. Moreover, also different frameworks for quantitative equational reasoning could be considered, for instance the one that has been introduced recently by Gavazzo and Di Florio [GD23].

5 Acknowledgements

This contribution was written within the framework of the SQUEE project, which is supported by a grant from the Austrian Science Fund (FWF) under project P 35530. The authors wish to express their thanks to the anonymous reviewers, whose comments helped to improve the paper's quality.

References

- [Bac+20] Giorgio Bacci, Radu Mardare, Prakash Panangaden, and Gordon Plotkin. “Quantitative Equational Reasoning”. In: *Foundations of Probabilistic Programming*. Ed. by Gilles Barthe, Joost-Pieter Katoen, and Alexandra Silva. Cambridge University Press, 2020, pp. 333–360.
- [BN98] Franz Baader and Tobias Nipkow. *Term rewriting and all that*. Cambridge University Press, Cambridge, 1998, pp. xii+301.
- [GD23] Francesco Gavazzo and Cecilia Di Florio. “Elements of Quantitative Rewriting”. In: *Proc. ACM Program. Lang.* 7.POPL (Jan. 2023).
- [JR15] Pascual Julián-Iranzo and Clemente Rubio-Manzano. “Proximity-based unification theory”. In: *Fuzzy Sets and Systems* 262 (2015), pp. 21–43.
- [KP19] Temur Kutsia and Cleo Pau. “Matching and Generalization Modulo Proximity and Tolerance Relations”. In: *Language, Logic, and Computation - 13th International Tbilisi Symposium, TbiLLC 2019, Batumi, Georgia, September 16-20, 2019, Revised Selected Papers*. Ed. by Aybüke Özgün and Yulia Zinova. Vol. 13206. Lecture Notes in Computer Science. Springer, 2019, pp. 323–342.
- [MPP16] Radu Mardare, Prakash Panangaden, and Gordon Plotkin. “Quantitative Algebraic Reasoning”. In: *Proceedings of the 31st Annual ACM/IEEE Symposium on Logic in Computer Science*. LICS '16. New York, NY, USA: Association for Computing Machinery, 2016, pp. 700–709.
- [Pau22] Cleo Pau. “Symbolic Techniques for Approximate Reasoning”. PhD thesis. RISC, JKU, 2022.

- [PK21] Cleo Pau and Temur Kutsia. “Proximity-Based Unification and Matching for Fully Fuzzy Signatures”. In: *30th IEEE International Conference on Fuzzy Systems, FUZZ-IEEE 2021, Luxembourg, July 11-14, 2021*. IEEE, 2021, pp. 1–6.
- [Ses02] Maria I. Sessa. “Approximate reasoning by similarity-based SLD resolution”. In: *Theoret. Comput. Sci.* 275.1-2 (2002), pp. 389–426.