

# The ABSURDIST algorithm for matching concept systems as a converging optimization method

April 26, 2004

The ABSURDIST algorithm introduced in [2] and further elaborated on in [1].

[Here, briefly Summarize what ABSURDIST is doing.]

In this report, we try to analyze ABSURDIST from the point of view commonly used for iterative methods. When an iterative numerical method ( $\vec{x}_{t+1} = f(\vec{x}_t)$ , with some initial starting point  $\vec{x}_0$ ) is proposed for solving a system of equations or an optimization problem, it is customary to analyze its behavior to answer these kinds of questions:

- Does the iterative process converge?
- What is the convergence point (or points)?
- Is the convergence global (the process converges to the known convergence point regardless of what the initial point  $\vec{x}_0$  is), or local (convergence to the stable point  $\vec{x}_*$  happens only if  $\vec{x}_0$  is located in a certain vicinity of  $\vec{x}_*$ )?
- Can we estimate the convergence rate? It is desirable if we can prove, for example, that  $\|\vec{x}_t - \vec{x}_*\| \leq \text{const} \cdot \|\vec{x}_t - \vec{x}_*\| \cdot r^t$ , with some convergence rate  $r < 1$ .

The analysis in this report presents partial answers to some of these questions, in the assumption of the constant *chi/beta* ratio throughout the iterative process.

In Section 1, we discuss how the general update formula in ABSURDIST works, and indicate the safe range of the values of the learning rate that should be used to ensure that the correspondence matrix elements stay within the prescribed range.

In Section 2, we show that ABSURDIST can be interpreted as a development on the classic steepest descent method for optimization of a certain “energy functional” on the set of matrix whose element value are subject to a number of constraints. We further restrict the range of the learning rate that should be used to ensure convergence.

In Section 3, we discuss a number of likely local convergence points of the ABSURDIST algorithm, and indicate the conditions under which local convergence to them happens.

In Section 4, the global maximum of the ABSURDIST energy functional is discussed, and interpreted in terms of the quality of the concept system permutation it generates.

To simplify calculations, we will assume that the size of both systems is the same,  $n$  nodes. (We use  $n$  instead of  $N$  for node number to avoid confusion with the net input matrix  $N$ ).

## 1 Update formula analysis

The ABSURDIST carries out an iterative process on an  $n \times n$  correspondence matrix  $C_t$ , whose element  $C_t(q, x)$  denotes the putative correspondence between the concepts  $A_q$  in system  $A$  and  $B_x$  in system  $B$  at the iteration step  $t$ . All elements of  $C$  are supposed to be in the range  $c_{\min} \leq C(q, x) \leq c_{\max}$ ; that is, the matrix  $C \in R^{n^2}$  is within the cube  $Q$  defined as

$$Q = \{C : (\forall q, x) \quad (c_{\min} \leq C(q, x) \leq c_{\max})\}.$$

At every iterative step, ABSURDIST computes the net input matrix  $N_t = N_t(C)$  with coefficients

$$N_t(A_q, B_x) = \alpha E(A_q, B_x) + \beta R_t(A_q, B_x) - \chi I_t(A_q, B_x), \quad (1)$$

where  $E$ ,  $R_t$ , and  $I_t$  are the external similarity, excitation, and inhibition, respectively. The external similarity coefficients describe our *a priori* beliefs about the similarity of concepts  $A_q$  and  $B_x$  in their respective concept systems, and are constant throughout the iteration process; meanwhile the excitation and inhibition at the step  $t$  are computed based on the correspondence matrix at this step:

$$R_t(A_q, B_x) = \frac{1}{n-1} \sum_{r \neq q} \sum_{y \neq x} S(\vec{a}_{qr}, \vec{b}_{xy}) C_t(A_r, B_y) \quad (2)$$

$$I_t(A_q, B_x) = \frac{1}{2(n-1)} \left( \sum_{r \neq q} C_t(A_r, B_x) + \sum_{y \neq x} C_t(A_q, B_y) \right) \quad (3)$$

Here  $S(\vec{a}_{qr}, \vec{b}_{xy})$  is the *edge similarity function* that measures the similarity between  $\vec{a}_{qr}$ , the bundle of relations existing in both directions in the concept pair  $(A_q, A_r)$  in system  $A$ , and  $\vec{b}_{xy}$ , the bundle of relations existing in the pair  $(B_x, B_y)$  in system  $B$ . As in [1], we assume that this is a symmetrized function, constructed as follows:

$$S(\vec{a}_{qr}, \vec{b}_{xy}) = S(\vec{a}_{rq}, \vec{b}_{yx}) = 1 - D(\vec{a}_{qr}, \vec{b}_{xy}), \quad (4)$$

with

$$D(\vec{a}_{qr}, \vec{b}_{xy}) = 1 - (D_d(\vec{a}_{qr}, \vec{b}_{xy}) + D_d(\vec{a}_{rq}, \vec{b}_{yx}))/2.$$

Here  $D_d(\vec{a}_{qr}, \vec{b}_{xy}) \in [0, 1]$  is the “directed” difference between the relations in the two bundles, taking into account only directed relations in the  $q \rightarrow r$  and  $x \rightarrow y$  directions, and undirected relations. It is 1 if the bundles being compared in the two systems are identical, and 0 if the bundles are absolutely different (each relation is either present with maximum weight in  $\vec{a}_{qr}$  and completely absent in  $\vec{b}_{xy}$ , or vice versa). Due to our definition, the values of the edge similarity function are always in the  $[0; 1]$  range.

Our edge similarity function is an extension of the “similarity of psychological distances” measure used in the original ABSURDIST.

At every step, the correspondence matrix is modified by the net input in a way that is intended to keep the  $C_{t+1}$  within  $Q$ :

$$C_{t+1} = C_t + LV_t, \quad (5)$$

where  $L > 0$  is the learning rate, and the update  $V_t$  is obtained by “damping” the net input in some way:

$$V_t = \text{Damp}(R_t). \quad (6)$$

The simplest way of “damping” the increment would be by truncating components of  $C_{t+1}$  to keep it within  $Q$ :

$$V_t(A_q, B_x) = \begin{cases} (c_{\max} - C_t(A_q, B_x))/L, & \text{if } C_t(A_q, B_x) + LN_t(A_q, B_x) > c_{\max}; \\ (c_{\min} - C_t(A_q, B_x))/L, & \text{if } C_t(A_q, B_x) + LN_t(A_q, B_x) < c_{\min}; \\ N_t(A_q, B_x), & \text{otherwise} \end{cases} \quad (7)$$

with learning rate  $L$ . However, the ABSURDIST algorithm in [2] uses a more complex damping scheme that results in a non-linear increment:

$$V_t(A_q, B_x) = \begin{cases} N_t(A_q, B_x) \cdot (c_{\max} - C_t(A_q, B_x)), & \text{if } N_t(A_q, B_x) \geq 0; \\ N_t(A_q, B_x) \cdot (C_t(A_q, B_x) - c_{\min}), & \text{if } N_t(A_q, B_x) < 0. \end{cases} \quad (8)$$

Intuitively, the difference between the two increment damping schemes is that with the simple linear increment, a positive or negative net input continuing over a sufficient number of steps will make the appropriate (upper or lower) bound in a finite number of steps; meanwhile with the ABSURDIST non-linear increment scheme, the value of  $C_t(q, x)$  will be approaching the bound exponentially, never actually reaching it. More on this in Section 3.

Under which conditions will the increment damping formula (8) ensure that, for any matrix  $C_t \in Q$ , the matrix  $C_{t+1}$  will stay within  $Q$  as well? To satisfy this condition, we can require that  $|LN_t(A_q, B_x)| \leq 1$  for all possible value of  $N_t(A_q, B_x)$ . That can be achieved if we choose the learning rate  $L$  so that

$$0 < L \leq \left( \max_{C \in Q} \|N(C)\|_\infty \right)^{-1}. \quad (9)$$

(The notation  $\|C\|_\infty$  is used for the “infinity norm” of a matrix  $C$ ,  $\|C\|_\infty = \max_{q,x} |C(q, x)|$ .)

The value of  $\|N(C)\|_\infty$  depends on the formula for the net input. Under (1,2,3) with non-negative  $\beta$  and  $\chi$ ,

$$E(A_q, B_x) - \chi \leq N(A_q, B_x) \leq E(A_q, B_x) + \beta(n-1) + \chi,$$

and therefore

$$\|N(C)\|_\infty \leq \|E\|_\infty + \beta(n-1) + \chi. \quad (10)$$

If all external similarity values  $E(A_q, B_x)$  are non-negative, a tighter bound obtains:

$$\|N(C)\|_\infty \leq \|E\|_\infty + \max(\beta(n-1), \chi)$$

Therefore, a constraint such as

$$0 < L \leq 1/\|E(C)\|_\infty \leq 1/(\|E\|_\infty + \beta(n-1) + \chi) \quad (11)$$

will guarantee that all values of the correspondence matrix stay within the  $[c_{\min}, c_{\max}]$  range, as long as the initial values are in that range. With  $0 \leq E(A_q, B_x) \leq$ , a sufficient condition on  $L$  is

$$0 < L \leq 1/\max(1 + \beta(n-1), \chi) \quad (12)$$

## 2 ABSURDIST as a converging optimization algorithm

**Assumptions.** In the rest of this report, we will assume, unless otherwise indicated, that:

- $\beta > 0$  (the excitation factor is always present, with a positive sign, in the net input).
- $\alpha \geq 0$ .
- $\chi \geq 0$ .
- The bounds on the elements of  $C$  are assumed  $c_{\min} = 0$  and  $c_{\max} = 1$ , as in [2]. The setting of  $c_{\min} = 0$  is necessary for some results, but the setting  $c_{\max} = 1$  is simply used for convenience of notation.
- The initial matrix  $C_0$  is chosen within cube  $Q$  (i.e., all its elements are in the  $[c_{\min}, c_{\max}]$  range).
- The learning rate  $L$  in (8) is chosen subject to the constraint (11), as to ensure that all  $C_t$  stay within cube  $Q$ .

**Convergence analysis.** In this section we will show that if the learning rate  $L$  is chosen sufficiently small, the ABSURDIST algorithm always converges, and at its convergence point is almost always a local maximum of a certain functional is reached.

In this section we will look at  $n \times n$  matrices, such as the correspondence matrix  $C$  or the net input matrix  $R$ , as at vectors from the  $n^2$ -dimensional Euclidean space  $R^{n^2}$ . For elements of this space we will use the standard dot product,

$$(C_1, C_2) = \sum_{q,x} C_1(q, x)C_2(q, x),$$

and the 2-norm,

$$\|C\|_2 = (C_1, C_2)^{1/2}.$$

We can interpret the formulae (1,2,3) for computing net input  $N \in R^{n^2}$  in terms of a linear operator  $\mathcal{A} : R^{n^2} \rightarrow R^{n^2}$ :

$$N(C) = \alpha E + \mathcal{A}C,$$

where

$$\mathcal{A}C = \beta R(C) - \chi I(C).$$

$\mathcal{A}$  can be represented as an  $n^2 \times n^2$  matrix, with its coefficients determined by the coefficients in the formulas for  $R(C)$  and  $I(C)$ . Due to the symmetry (4) of the edge similarity function with respect to edge direction,  $\mathcal{A}$  is a self-adjoint (symmetric) operator, i.e.

$$(u, \mathcal{A}v) = (v, \mathcal{A}u) \quad (\forall u, v \in R^{n^2}).$$

For the linear operator  $\mathcal{A}$ , we can use the standard operator norm induced on  $R^{n^2} \rightarrow R^{n^2}$  by the vector 2-norm in  $R^{n^2}$ , that is

$$\|\mathcal{A}\| = \max_{C \neq 0} \frac{\|\mathcal{A}C\|_2}{\|C\|_2}.$$

Let us now introduce the *energy functional*  $K : R^{n^2} \rightarrow R$ :

$$K(C) = \frac{1}{2}(C, \mathcal{A}C) + \alpha(E, C). \quad (13)$$

In terms of individual matrix elements, this functional can be represented as

$$\begin{aligned} K(C) = & \frac{\beta}{2(n-1)} \sum_{q,r:r \neq q} \sum_{x,y:y \neq x} S(\vec{a}_{qr}, \vec{b}_{xy}) C(A_q, B_x) C(A_r, B_y) - \\ & \frac{\chi}{4(n-1)} \left( \sum_{q,r:r \neq q} \sum_x C(A_q, B_x) C(A_r, B_x) + \sum_{x,y:x \neq y} \sum_q C(A_q, B_x) C(A_q, B_y) \right) + \\ & \alpha \sum_q \sum_x E(A_q, B_x) C(A_q, B_x). \end{aligned}$$

The energy functional  $K(C)$  is defined in such a way that its gradient is  $R(C)$ . Therefore, if the net input  $N_t$  were directly added to  $C_t$  without damping (i.e., if the damping function in (6) were an identity function,  $\text{Damp}(N) = N$ ), then the ABSURDIST algorithm would become the classic steepest descent method [GIVE SOME STANDARD REFERENCE] for the minimization of  $-K(C)$ . We will show that ABSURDIST, which uses damped increments, has similar optimization properties.

As the following proposition shows, if the learning rate  $L$  is sufficiently low, the energy functional of  $C_t$  will be monotonically increasing during ABSURDIST iteration; the increase may only stop if iterative process has converged after a finite number of steps.

**Proposition 1.** *Consider the iterative process*

$$\begin{aligned} N_t &= \alpha E + \mathcal{A}C_t, \\ V_t &= \text{Damp}(N_t, C_t), \\ C_{t+1} &= C_t + LV_t, \end{aligned}$$

in the linear space  $R^m$ , where  $\mathcal{A}$  is a self-adjoint operator, the learning rate  $L$  is within the range

$$0 < L < 2/\|\mathcal{A}\|, \quad (14)$$

and the continuous function  $\text{Damp}(\cdot)$  has the following “componentwise damping property”: it preserves signs, but does not increase the absolute value of each component of its first argument. That is, for each

coordinate  $i$ ,

$$\begin{aligned} 0 < \text{Damp}(N, C)^i &\leq N^i && \text{if } N^i > 0 \text{ and } C_i < c_{\max}, \\ \text{Damp}(N, C)^i &= 0 && \text{if } N^i > 0 \text{ and } C_i = c_{\max}, \\ 0 < \text{Damp}(N, C)^i &\geq N^i && \text{if } N^i < 0 \text{ and } C_i < c_{\max}, \\ \text{Damp}(N, C)^i &= 0 && \text{if } N^i < 0 \text{ and } C_i = c_{\min}. \end{aligned}$$

In this iterative process, the energy increases at every step as long as  $C_t$  has not converged yet.

*Proof.* By construction and due to the self-adjointness of  $\mathcal{A}$ ,

$$K(C_{t+1}) - K(C_t) = \frac{1}{2}L^2(V_t, \mathcal{A}V_t) + L(V_t, \mathcal{A}C) + (V_t, E) = L\left(\frac{L}{2}(V_t, \mathcal{A}V_t) + (V_t, N_t)\right).$$

Due to the limitation on  $L$ , we can bound the absolute value of the first term with

$$\left|\frac{L}{2}(V_t, \mathcal{A}V_t)\right| \leq \frac{L\|\mathcal{A}\|}{2}\|V_t\|_2^2 \leq \|V_t\|_2^2.$$

Since the absolute values of the components of the vector  $V_t = \text{Damp}(N_t, C_t)$  are no greater than those of  $N_t$ ,

$$0 \leq \|V_t\|_2^2 \leq (V_t, N_t) \leq \|N_t\|_2^2.$$

Therefore,  $|\frac{L}{2}(V_t, \mathcal{A}V_t)| \leq \frac{L\|\mathcal{A}\|}{2}\|V_t\|_2^2 \leq (V_t, N_t)$ , and  $\frac{L}{2}(V_t, \mathcal{A}V_t) + (V_t, N_t) \geq 0$ , with the equality reached only when  $V_t = 0$ . Hence,  $K(C_{t+1}) > K(C_t)$  as long as  $V_t \neq 0$ . If  $V_t = 0$ , the sequence has converged ( $C_m = C_t$  for any  $m \geq t$ ), and  $K(C_t)$  will of course stay constant from this point on.  $\square$

Note that the the specific formula for the damping function does not matter for the above proposition, as long as this function possesses the ‘‘damping property’’. Therefore, both the ABSURDIST algorithm, which uses the ‘‘soft’’ damping given by (8), and a similar algorithm with ‘‘hard’’ damping (7) satisfy Proposition 1.

To estimate the bound in the restriction restriction (14) on learning rate  $L$ , we note that for the self-adjoint matrix  $\mathcal{A}$ , its matrix norm  $\|\mathcal{A}\|$  induced by the 2-norm on  $R^{n^2}$  does not exceed the matrix norm induced by the infinity-norm,  $\max_{C:\|C\|_\infty=1} \|R(C) + I(C)\|_\infty \leq \beta(n-1) + 2 * \chi$ . Therefore, bounding the learning rate by

$$0 < L < 2/(\beta(n-1) + 2 * \chi) \leq 2/\|\mathcal{A}\| \tag{15}$$

is sufficient to ensure monotonic increase of energy. This restriction is similar, but not identical, to restriction (11). In practice, it may be desirable to set  $L$  so that it would satisfy both limits, well inside both of the permitted ranges (11) and 15).

Based on one’s general idea of the behavior of linear and quadratic functions, it seems intuitively obvious that during the ABSURDIST iterative process not only does the energy functional  $K(C_t)$  converge to  $K_*$ , but the actual correspondence matrix  $C_t$  converges to some  $C_*$ . Unfortunately, we don’t have a short cogent proof of this. Instead, we present this result as a conjecture, with a sketch of a likely correct, but unwieldy, proof.

**Conjecture 1.** *When the learning rate  $L$  is  $0 < L < 2/\|\mathcal{A}\|$ , the sequence of correspondence matrices  $\{C_0, C_1, \dots\} \subset Q$  computed by the iterative process described in Proposition 1 converges to some matrix  $C_* \in Q$ .*

*Sketch of a proof.*

(1) By Proposition 1,  $K(K_t)$  is monotonically increasing. Since  $K(\cdot)$  is a continuous function on a compact set  $Q$ , it is bounded; therefore, there is  $K_*$  such that  $\lim_{t \rightarrow \infty} K(K_t) = K_*$

(2) Due to the compactness of cube  $Q$ , the sequence  $\{C_t\}_{t=1,2,\dots}$  contains a convergent subsequence  $\{C_{k_j}\}_{j=1,2,\dots}$ . We will designate its limit with  $\lim_{j \rightarrow \infty} C_{k_j} = C_*$ . It is easy to see that  $K(C_*) = K_*$ .

(3) It can be shown that if the subsequence limit  $C_*$  is a strict local energy maximum on  $Q$  (that is, if  $(\exists \varepsilon > 0) (\forall C : C \in Q \wedge \|C - C_*\| < \varepsilon) (K(C) < K(C_*))$ ), then the entire sequence converges to  $C_*$ :  $\lim t \rightarrow \infty C_t = C_*$ .

(4) We will now show that  $V_t$ , the increment to  $C_t$ , converges to 0. The proof is as follows: as in Proposition 1,

$$K(C_{t+1}) - K(C_t) = L \left( \frac{L}{2} (V_t, \mathcal{A}V_t) + (V_t, N_t) \right) > L \left( \left(1 - \frac{L\|\mathcal{A}\|}{2}\right) \|V_t\|_2^2 \right),$$

and therefore

$$\|V_t\|_2 < \sqrt{\frac{K(C_{t+1}) - K(C_t)}{L(1 - L\|\mathcal{A}\|/2)}}.$$

Since energy is bounded and monotonically increasing,  $K(C_{t+1}) - K(C_t) \rightarrow 0$ , and so does  $\|V_t\|_2$ .

(5) It can be shown that  $(\forall \varepsilon > 0) (\exists T) (\forall t > T) (\forall C \in Q) (N_t, C_t - C) > -\varepsilon$ . In other words, if  $C_*$  is the limit of a subsequence, then either  $C_*$  is located inside cube  $Q$  and  $N(C_*) = 0$ , or it is located on the boundary of  $Q$  and the only non-zero components of  $N(C_*)$  are those that are directed “outward” of the cube  $Q$ . (Give forward ref here).

(6) If  $N(C_*) = 0$ , then for any  $C$ ,  $K(C_*) - K(C) = \frac{1}{2}(C - C_*, \mathcal{A}(C - C_*)) = (N(C), \mathcal{A}^+N(C))$ , where  $\mathcal{A}^+$  is the pseudo-inverse of  $\mathcal{A}$ .

(7) This is the only complicated, and not strictly worked out, part of the proof. We start with considering the possibilities for the point  $C_*$  located inside  $Q$ . There are three:

- (i)(a):  $\mathcal{A}$  is negative definite, and on  $C_* = -\mathcal{A}^{-1}E$  the strict global maximum of  $K(\cdot)$  is reached. In this case clause (3) immediately applies. Moreover, it is also possible to show, by using clause (6), that  $\sum_{t=0}^{\infty} \|V_t\|_2$  is finite.
- (i)(b):  $\mathcal{A}$  is negative semidefinite (singular) matrix, with  $E$  orthogonal to  $\ker \mathcal{A}$ . In this case the energy maximum is reached on any point  $C'$  such that  $C' + \mathcal{A}^+E \in \ker \mathcal{A}$ . In this case any such  $C'$  will be a non-strict global maximum, and  $C_*$  will be one of them. Our task here is to show that the sequence  $\{C_t\}_{t=1,2,\dots}$  still converges to a single point. We believe this can be proven by decomposing each increment  $V_t$  into two components:  $V_t^{\parallel} \in \ker \mathcal{A}$ , and  $V_t^{\perp}$ , orthogonal to  $\ker \mathcal{A}$ . Using the continuity of the damping function, and the fact that each  $N_t$  is orthogonal to  $\ker \mathcal{A}$ , it is possible to prove that the  $\|V_t^{\parallel}\|_2 \leq \zeta \|V_t^{\perp}\|_2$  for certain constant  $\zeta$ . Further on, we believe that it is possible to show that  $\sum_{t=0}^{\infty} \|V_t^{\perp}\|_2$  is finite (in a way similar to (i)(a)), and therefore,  $\sum_{t=0}^{\infty} \|V_t^{\parallel}\|_2$  and  $\sum_{t=0}^{\infty} \|V_t\|_2$  are finite as well, which will prove the convergence of  $C_t$ .
- (i)(c): A theoretically possible case is that of a “saddle point”, which may occur with a  $\mathcal{A}$  that has both positive and negative eigenvalues. Since convergence to such a point is unstable (a small variation in  $C_0$ , or a small error introduced in any  $C_t$ , may destroy the convergence), an occurrence of such an even in a practical computation is exceedingly unlikely. If  $C_*$  is such a saddle point, it appears that it is still possible to prove that it is the convergence point for the entire sequence by using some the projections to the subspace based on the negative- $\lambda$  eigenvectors of  $\mathcal{A}$ .

We will now consider the second possibility, when the subsequence convergence point  $C_*$  sits on the boundary of the cube  $Q$ , i.e.  $C_*^i = c_{\min}$  or  $C_*^i = c_{\max}$  for some coordinates  $i$ . Consider projection operators for two subspaces:  $\mathcal{P}_1$ , for the coordinates for which  $C_*^i = c_{\min}$  or  $C_*^i = c_{\max}$ , and  $\mathcal{P}_2$ , for the remaining coordinates. By (5), we see that  $N(C_*)$  lies in the first space. In that case it appears that one can “split” the coordinates of  $C$ , proving the convergence of  $\mathcal{P}_1 C_t$  by a technique similar in Proposition 2, below, while applying one of the cases (i)(a-c) to the convergence of  $\mathcal{P}_2 C_t$ .  $\square$

### 3 Local convergence points

We have shown in the previous section that ABSURDIST with a low enough learning rate  $L$  always converges to some correspondence matrix  $C_*$ , and the convergence point  $C_*$  almost always is a (strict or non-strict) local maximum of the energy functional  $K(C)$  on cube  $Q$  in the space of  $n \times n$  matrices  $R^{n^2}$ . This functional

may, of course, have more than one local maximum, and every one of them is a convergence point for an iterative sequence started at some point in  $Q$ . (As a trivial example, one can see that an iterative sequence started at a local-maximum  $C_*$  will simply stay at this  $C_*$ , that is will immediately “converge” to  $C_*$ ).

The properties of all convergence points  $C_*$  are given by clause (3) in the proof sketch of Conjecture 1, and the possibilities are delineated in clause (7) of the same sketch. To visualize the possibilities, one can imagine a quadratic polynomial on a cube in a 2-D or 3-D space, and figure where it may have its maxima: either a single point inside the cube (when the operator  $\mathcal{A}$  is negative definite, and  $-\mathcal{A}^{-1}E \in Q$ ), or a (hyper)plane traversing the cube (when the operator  $\mathcal{A}$  is negative semi-definite, and  $-\mathcal{A}^+E + \ker\mathcal{A} \cap Q \neq \emptyset$ ), or a point on one of the faces or edges of the cube, with the vector  $N(C_*)$  directed outward from the cube.

### 3.1 Conditions of convergence to a matrix of $c_{\min}$ and $c_{\max}$ values

A particularly simple case of a ABSURDIST local convergence is to a matrix  $C_*$  that’s in a corner of the cube  $Q$  – that is, a matrix that consists only of values  $c_{\min}$  and  $c_{\max}$ . We will first observe that a simple sufficient condition exists for ABSURDIST to converge to such a matrix.

**Proposition 2.** *If the correspondence matrix  $C_*$  consists only of values  $c_{\min}$  and  $c_{\max}$ , and the elements of the net input matrix  $N_* = N(C_*)$  have the following signs:*

$$\begin{aligned} N_*(q, x) < 0 & \quad \text{if } C_*(q, x) = c_{\min}; \\ N_*(q, x) > 0 & \quad \text{if } C_*(q, x) = c_{\max}; \end{aligned}$$

*then  $C_*$  is a local convergence point for any ABSURDIST iterative process starting anywhere within the intersection of the cube  $Q$  and a certain vicinity of  $C_*$ .*

*Proof.* Let  $a = \min_{q,x} |N_*(q, x)| > 0$ . Since  $N$  is a continuous function of  $C$ , for any positive  $a' < a$  there is such an  $\varepsilon > 0$  that for any  $C$  within the vicinity  $V_\varepsilon(C_*) = \{C : (C \in Q) \wedge \|C - C_*\|_\infty < \varepsilon\}$ , we will have  $\min_{q,x} |N(C)(q, x)| > a'$ . Pick any such  $a'$ , say  $a' = a/2$ , and the appropriate  $\varepsilon$ .

According to the ABSURDIST update scheme, if  $N(q, x) < 0$ , then  $C_{t+1}(q, x) = (1 + LN(q, x))C_t$ ; if  $N(q, x) > 0$ , then  $(1 - C_{t+1}(q, x)) = (1 - LN(q, x))(1 - C_t)$ . Therefore, for any  $C_t \in V_\varepsilon(C_*)$ , we will have  $\|C - C_*\|_\infty \leq (1 - La')\|C - C_*\|_\infty$ . This means that for any starting point  $C_0 \in V_\varepsilon(C_*)$ , the ABSURDIST iterative process converges exponentially to  $C_*$ :

$$\|C_t - C_*\|_\infty \leq r^t \|C_0 - C_*\|_\infty, \tag{16}$$

with the convergence rate  $r = 1 - La'$ . □

A nearly converse can be proven as well: If the correspondence matrix  $C_*$  consists of 0s and 1s, and some of the net inputs has a “wrong” sign: either  $N_*(q, x) > 0$  with  $C_*(q, x) = c_{\min}$ , or  $N_*(q, x) < 0$  with  $C_*(q, x) = c_{\max}$ , then ABSURDIST cannot converge to  $C_*$ .

As a side note, we can notice that if the linear increment scheme (7) were used for updating the correspondence matrix at each step, the iterative process starting within the same vicinity  $V_\varepsilon$  would still converge, but now in a finite number of steps (no more than  $\|C_0 - C_*\|_\infty/a'$ ), rather than exponentially.

In the rest of this section, we will apply Proposition 2 to several potential convergence points of interest.

### 3.2 The all- $c_{\max}$ matrix

We will show here that for some wide classes of concept systems and a sufficiently low  $\chi/\beta$  ratio, the matrix  $C_* = c_{\max}ee^T$ , with all  $C(q, x) = c_{\max}$ , will be an ABSURDIST convergence point. ( $e \in R^n$  is a vector of all 1s).

**Proposition 3.** *The correspondence matrix  $C_* = c_{\max}ee^T$  is convergence point for any ABSURDIST iterative process started within a certain vicinity of  $C_*$  if*

$$\frac{\chi}{\beta} < H(A, B), \tag{17}$$

where

$$H(A, B) = \frac{1}{n-1} \min_{A_q \in A, B_x \in B} \sum_{r \neq q} \sum_{y \neq x} S(\vec{a}_{qr}, \vec{b}_{xy}) \quad (18)$$

*Proof.* The net inputs  $N_* = N(C_*)$  at  $C_*$  are

$$N_*(q, x) = \alpha E(q, x) + \beta \frac{1}{n-1} \sum_{r \neq q} \sum_{y \neq x} S(\vec{a}_{qr}, \vec{b}_{xy}) - \chi.$$

We have assumed that external similarities are non-negative; therefore, if  $\chi/\beta < H(A, B)$ , then all  $N_*(q, x)$  will be positive. By Proposition 2,  $C_*$  is the convergence point for any ABSURDIST iterative process starting within a certain vicinity of  $C_*$ .  $\square$

While  $H(A, B)$  is not exactly a measure of concept system similarity, it measures something related: how similar the “least similar” concepts in  $A$  and  $B$  are, with respect to their “outlook” to the rest of the concept system. Since the edge similarity function  $S$  is non-negative,  $H(A, B)$  is non-negative too. It will be 0 only if some concepts  $A_q$  in system  $A$  are very “different” from some concept  $B_x$  in system  $B$ —that is, there is not a single pair of concepts ( $A_r \in A, B_y \in B$ ) such that  $S(\vec{a}_{qr}, \vec{b}_{xy}) > 0$ .

The likelihood of that happening depends on what kind of relations there are in the system, as well as on the degrees of the system graph.

For example, if no concept in either system has direct relations with all other nodes, then we can show that  $H(A, B) > 0$ . Indeed, in these pairs of systems for any  $A_q \in A$  there is an  $A_r$  not connected to  $A_q$ , and for any  $B_x \in B$  there is a  $B_y$  not connected to  $A_x$ ; thus, for these pairs of concepts,  $S(\vec{a}_{qr}, \vec{b}_{xy}) = 1$ .

For a concept system with only one unweighted, undirected relation (representable by an unweighted undirected graph),  $H(A, B)$  evaluates to

$$H(A, B) = \min_{A_q \in A, B_x \in B} \left( n - 1 - \deg(A_q) - \deg(B_x) + \frac{2\deg(A_q)\deg(B_x)}{n-1} \right), \quad (19)$$

where  $\deg(A_q)$  is the degree of node  $A_q$  (the number of concepts in  $A$  to which  $A_q$  is directly related to). Consider a rather common case of sparse graphs where each node is linked to only a few nodes (less than half of all nodes in the system):

$$\max_{A_q \in A} \deg(A_q) = D_A < (n-1)/2, \quad \max_{B_x \in B} \deg(B_x) = D_B < (n-1)/2.$$

For such a system,  $H(A, B)$  in (19) can be evaluated as

$$H(A, B) = n - 1 - D_A - D_B + 2D_A D_B / (n - 1) > 0.$$

Thus, on such a pair of concept systems, the matrix of all  $c_{\max}$  will be an ABSURDIST convergence point if

$$\chi/\beta < n - 1 - D_A - D_B + 2D_A D_B / (n - 1).$$

On the other hand, in a pair of systems with several types of relations, it can be quite easy for  $H(A, B) = 0$ , in which case ABSURDIST is no danger of converging to  $c_{\max} e e^T$ , unless widespread external similarity makes it to.

### 3.3 A matrix with a block of $c_{\max}$ values

Even if  $H(A, B) = 0$ , and ABSURDIST does not converge to a matrix of all  $c_{\max}$ , it may be the case that a matrix with a block of  $c_{\max}$  values is a convergence point.

**Proposition 4.** *Suppose the concept systems  $A$  and  $B$  contain subsets  $A' \in A$  and  $B' \in B$  such that for each  $q \in A'$  and for each  $x \in B'$  there are such  $r \in A'$  and for each  $y \in B'$  that  $S(\vec{a}_{qr}, \vec{b}_{xy}) > 0$ . Suppose that  $\chi/\beta < H(A', B')$ , where*

$$H(A', B') = \min_{A_q \in A', B_x \in B'} \frac{\sum_{r \neq q} \sum_{y \neq x} S(\vec{a}_{qr}, \vec{b}_{xy})}{n-1} > 0. \quad (20)$$

Then there is such a  $\varepsilon > 0$  that any ABSURDIST iterative process starting at a point where  $(\forall q \in A') (\forall x \in B') (C(q, x) > c_{\max} - \varepsilon)$ , will converge to a certain matrix  $C_*$  with  $(C(q, x) = c_{\max})$  for all  $q \in A'$  and  $x \in B'$ .

*Proof.* We will notice first that  $H(A', B') > 0$ . Similarly to Proposition 4, it can be shown that if  $\chi/\beta \leq H(A', B')$ , then for some sufficiently small  $\varepsilon > 0$ , for any correspondence matrix  $C \in Q$  such that  $(\forall q \in A') (\forall x \in B') (C(q, x) > c_{\max} - \varepsilon)$ , we will have  $(\forall q \in A') (\forall x \in B') (N(C)(q, x) > 0)$ . It follows therefore that during an ABSURDIST iterative process started anywhere in that region all the matrix components  $C_t(q, x)$  with  $q \in A'$  and  $x \in B'$  can only increase, until  $C_t$  converges to a point  $C_*$  with  $C_t(q, x) = c_{\max}$  for all  $q \in A'$  and  $x \in B'$ .  $\square$

Subsystem pairs  $(A', B')$  with the property required by Proposition 4 are quite common. For example if the two concept systems include a symmetric relation  $r(\cdot, \cdot)$ , an  $A'$  will be formed by the set of concepts  $A_q \in A$  such that there is an  $A_r \in A$  with  $r(A_q, A_r) > 0$ , and a  $B'$  will be formed by the set of concepts  $B_x \in A$  such that there is a  $B_y \in B$  with  $r(B_x, B_y) > 0$ .

### 3.4 Permutation matrices

In this section we will need to assume that  $c_{\min} = 0$ , as in [2]; for convenience, we will also assume that  $c_{\min} = 1$ . As usual, we assume that external similarities are non-negative.

We will look at permutation matrices. A permutation  $P$  is a bijective function from  $\{1, \dots, n\}$  to  $\{1, \dots, n\}$ , which maps  $q$  to  $P(q)$ , and thus can be thought of as mapping a concept  $A_q$  in  $A$  to the concept  $B_{P(q)}$  in  $B$ . We will use the notation  $P(A)$  to refer to the ‘‘permuted’’ system  $A$ , i.e. the concept system in which the same relations exist between the  $P(q)$ -th and  $P(r)$ -th concepts as do between  $A_q$  and  $A_r$  in  $A$ . (That is,  $\vec{a}'_{P(q), P(r)} = \vec{a}_{qr}$ , where  $\vec{a}'_{P(q), P(r)}$  denotes the bundle of relations existing in  $P(A)$  between  $P(a)_q$  and  $P(A)_r$ ).

In terms of correspondence matrices that ABSURDIST computes, permutation  $P$  can be described by the permutation matrix  $C_P$ . This is a matrix where  $C_P(q, x) = \delta_{P(q), x}$ ; that is,  $C_P(q, P(q)) = 1$  for each  $A_q \in A$ , and all other matrix elements are zeros.

We will now formulate a sufficient condition for ABSURDIST to converge to a particular permutation matrix:

**Proposition 5.** *If  $P$  is a permutation of  $\{1, \dots, n\}$ , the permutation matrix  $C_P$  will be a convergence point for ABSURDIST iterations started within a certain vicinity of  $C_P$  within cube  $Q$  if the following two conditions hold:*

1. For each  $q$ , either  $E(A_q, B_{P(q)}) > 0$ , or there is at least one  $r \neq q$  such that  $S(\vec{a}_{qr}, \vec{b}_{P(q), P(r)}) > 0$ ;
- 2.

$$\chi > \max_{q, x: x \neq P(q)} \left( \beta \sum_{r: r \neq q, P(r) \neq x} S(\vec{a}_{qr}, \vec{b}_{x, P(r)}) + \alpha(n-1)E(A_q, B_x) \right). \quad (21)$$

*Proof.* Since nothing specific is claimed about ordering of concepts in  $A$  and  $B$ , we can simplify our notation, without a loss of generality, by assuming that  $P$  is an identity permutation. In this case  $C(A_q, B_x) = \delta_{qx}$ , and the net input  $N(C)$  can be expressed as follows:

$$N(A_q, B_x) = \alpha E(A_q, B_x) + \frac{1}{n-1} \left( \beta \sum_{r: r \notin \{q, x\}} S(\vec{a}_{qr}, \vec{b}_{xr}) - \chi(1 - \delta_{qx}) \right).$$

This gives diagonal elements

$$N(A_q, B_q) = \alpha E(A_q, B_q) + \frac{1}{n-1} \left( \beta \sum_{r: r \neq q} S(\vec{a}_{qr}, \vec{b}_{qr}) \right).$$

and off-diagonal elements ( $q \neq x$ )

$$N(A_q, B_x) = \alpha E(A_q, B_x) + \frac{1}{n-1} \left( \beta \sum_{r:r \neq q, P(r) \neq x} S(\vec{a}_{qr}, \vec{b}_{xr}) - \chi \right).$$

The first and second conditions of this proposition ensure, respectively, that the diagonal elements of the net input matrix are positive, and the off-diagonal ones are negative. By Proposition 2, local convergence to  $C_P$  immediately follows.  $\square$

A nearly converse can be shown as well: if for any  $(q, x)$  pair  $\chi < \beta \sum_{r:r \neq q, P(r) \neq x} S(\vec{a}_{qr}, \vec{b}_{x,P(r)}) + \alpha(n-1)E(A_q, B_x)$ , then  $C_P$  won't be a local convergence point.

What is the meaning of the two conditions of Proposition 5?

The first condition simply means that the permutation maps each concept  $A_q$  of  $A$  to a somewhat similar concept of  $B$ . The necessary degree of similarity between  $A_q$  and  $B_{P(q)}$  is provided either by presence of any positive amount of external similarity between the two concepts, or by existence of at least one other concept  $A_r$  such that the relation bundles  $\vec{a}_{qr}$  and  $\vec{b}_{P(q),P(r)}$  are not completely dissimilar.

The second condition sets the minimum  $\chi$  guaranteeing that  $C_P$  will be a local convergence point, provided the first condition holds. One can see that if

$$\chi > \alpha \|E\|_\infty + \beta(n-2), \quad (22)$$

this second condition will be always satisfied.

If there is no external similarity, inequality (21) simplifies to

$$\chi/\beta > F(A, B, P), \quad (23)$$

with

$$F(A, B, P) = \max_{q,x:x \neq P(q)} \Theta(P(q), P(A), x, B),$$

where

$$\Theta(q', A', x, B) = \sum_{y:y \neq q', y \neq x} S(\vec{a}'_{q'y}, \vec{b}_{xy}).$$

One can view  $\Theta(P(q), P(A), x, B)$  as a function that compares the “view” from the concept  $A_q$  to the rest of system  $A$  with the “view” from  $x$  to  $B$ , taking the permutation into account. The more similar the “view”, the higher is  $\Theta$ . Its maximum possible value of  $\Theta$  is  $n-2$ ; it is reached if the permutation exactly aligns the concepts of  $A$  that are connected to  $q$  to the concepts of  $B$  that are connected to  $x$ , and each bundle of relations between  $A_q$  and another concept  $A_r$  in  $A$  is exactly the same as the those between  $x$  and the concept in  $B$  that is aligned with  $A_r$ .

**Example 1.** There are many situations when both conditions hold. Consider, for example, the case of two systems  $A$  and  $B$  where each concept is directly related in any way to fewer than  $(n-1)/2$  concepts. For any permutation  $P$ , for any  $q$  there will exist such an  $r$  that both  $\vec{a}_{qr}$  and  $\vec{b}_{P(q),P(r)}$  are empty bundles, and therefore  $S(\vec{a}_{qr}, \vec{b}_{P(q),P(r)}) = 1$ , thus satisfying the first condition. Thus if  $\chi$  is high enough to satisfy (22), **each** one of the  $2^n$  possible permutation matrices will be a local convergence point for ABSURDIST on  $(A, B)$ .

**Example 2.** An interesting case for ABSURDIST is matching two systems  $A$  and  $B$  that are isomorphous, i.e. can be matched exactly. Let's look first at the “correct” permutation  $P_*$ —the one that results in an exact match. For this permutation, the first condition is satisfied; as long as  $\chi$  satisfies (21), the second condition is satisfied as well, and  $C_{P_*}$  is a local convergence point.

However, as the previous example demonstrated, other permutation matrices (even all of them!) may be local convergence points as well, if the systems are sufficiently sparse and  $\chi$  is sufficiently high.

**Example 3.** May it be possible to set  $\chi/\beta$  sufficiently high to guarantee at least local convergence to the “correct” permutation matrix, but not too high to cause convergence to other, “incorrect” permutation matrices? It turns out not to be universally possible. Consider ABSURDIST trying to match of two isomorphous systems that can be exactly matched by permutation  $P$ , in the absence of external similarity.

Suppose that system  $B$  is symmetric with respect to swapping concepts  $x_1$  and  $x_2$ . In this case  $F(A, B, P) = \Theta(x_1, P(A), x_2, B) = n - 2$ , which is as high as  $F(A, B, P)$  can be. Therefore any  $\chi/\beta$  high enough to ensure local convergence to  $P$ , will also ensure local convergence to the permutation matrices of all other permutations satisfying Condition 1 of Proposition 5. As it has been noted above (Example 1), on some system classes this includes all permutations.

**Effect of external similarity.** As seen from inequality (21), the presence of positive external similarity may make it harder for ABSURDIST to locally converge to permutations “contradicted” by the external similarity, and easier to converge to the permutations that agree with the external similarity. For example, it can be used to break the symmetry described in Example 3 above. However, depending on a particular situation, a large number of external similarity coefficients may need to be provided to ensure convergence to the “right” permutation matrix.

## 4 Global maximum

It is natural to ask what the global maximum of the the energy functional (13) on cube  $Q$  is. In particular, if the functional is reached on a permutation matrix, is that permutation the “best” in some easily understandable way?

We will first consider values of the energy functional on all permutation matrices.

**Proposition 6.** *For any permutation  $P$  with permutation matrix  $C_P$ , the value of the energy functional  $K(C_P)$  on the permutation matrix  $C_P$  is*

$$K(C_P) = \alpha \sum_q E(q, P(q)) + \beta \sum_{q,r \in A, q \neq r} S(\vec{a}_{qr}, \vec{b}_{xy}) = \alpha \sum_q E(q, P(q)) + \beta(n - 2 * \mu(P(A), B)), \quad (24)$$

where the relations mismatch measure  $\mu(A, B)$  is

$$\mu(A, B) = \frac{1}{2} \sum_{q,r \in A, q \neq r} \sum_{x,y \in B, x \neq y} D(\vec{a}_{qr}, \vec{a}_{xy}).$$

*Proof.* Obtains from substituting a permutation matrix into the definition of the energy functional.  $\square$

The relations mismatch measure  $\mu(P(A), B)$  has a very similar interpretation: it is a sum that includes, for each of the  $n(n - 1)/2$  “potential edges” in  $B$ , the difference between this edge and its counterpart in  $A$  aligned to it by the permutation  $P$ . In the case of a concept system with just one symmetric unweighted relation the relations mismatch measure is simply the count of positions where the edges are misaligned, i.e. an edge is present in one graph but is absent in the corresponding position in the other.

From Proposition 6 it immediately follows that, in the absence of the external similarity, the permutations that out of all  $2^n$  permutation matrices the highest value of  $K(C_P)$  is reached on those permutations that are the best in the sense that they minimize the relations mismatch measure. In particular, if the two graphs are isomorphous, the highest-energy permutations, with  $K(C_P) = n$ , are those that deliver exact match of the two systems. If graphs are not isomorphous, the highest-energy permutation is the one that minimizes the edge mismatch. This minimal edge mismatch,  $\min_P \mu(P(A), B)$ , is rather similar to the edit distance measure used elsewhere [4, pp. 43-46], except that Messmer’s allowed operations include not only creation/destruction of edges (adding and removing relations between concepts) but also creation and removal of graph vertices (concepts).

If external similarity is present, permutations “contradicting” the external similarity (i.e., those where  $P(q) \neq x$  for some  $(q, x)$  pairs for which non-zero external similarity  $E(q, x) > 0$  is postulated) are penalized in (24). Therefore, if the permutation minimizing the edge mismatch function “contradicts” the external similarity, the some other permutation, achieving the best tradeoff between edge mismatch and external similarity, may turn out to be the energy maximizer.

What about the global maximum of energy on the entire cube  $Q$ , not just on permutation matrices? Its location generally depends on the parameters  $\alpha$ ,  $\beta$ , and  $\chi$ . Due to non-negativity of the edge similarity function  $S$  and the external similarity  $E$ , it is obvious that if  $\chi = 0$  (no inhibition), then the maximum of

$K(C)$  is reached on the matrix composed of all 1s,  $C = ee^T$ . As we increase  $\chi$ , and the contribution of the inhibition term proportionally increases, this matrix will cease to be the global (or even local) maximum. Although we don't have a proof for it, we conjecture that if  $\chi$  is sufficiently high, the global maximum will be reached at one of the permutation matrices. This conjecture is based on the following observations:

- The inhibition term  $(C, IC)$  is zero only on matrices in which all rows are mutually orthogonal, and so are all columns. The only matrices in  $Q$  with this property are those in  $\mathcal{P}$ : the set of matrix whose non-zeros are distributed in the same patterns as in a permutation matrix, i.e. at most non-zero per each row and per each column. Therefore  $K(C) \leq 0$  on all  $C \in \mathcal{P}$ , regardless of  $\chi$ .
- For any matrix  $C \in Q \setminus \mathcal{P}$ , there is a certain  $\chi(C)$  such that for any  $\chi > \chi(C)$  such that  $K(C) < 0$ . Therefore even if such a matrix is a global maximum point of  $K$  at some  $\chi$ , it will cease to be such as  $\chi$  increases beyond a certain value.
- For any  $C \in \mathcal{P}$ , we can construct a permutation matrix  $P$  which has a 1 in every position where  $C$  has a non-zero. By non-negativity of the edge similarity function,  $K(P) \geq K(C)$ . Therefore, even if the global maximum of  $K$  is reached on a  $C \in \mathcal{P}$  that is not a permutation matrix (which, in practice, may only happen in some “hard-to-match” pairs of systems, for which Condition 1 of Proposition 5 is not satisfied in any permutation), then the same value is reached on a permutation matrix as well.

## 5 Conclusions

In this report, we have presented an elementary analysis of the convergence of the ABSURDIST algorithm with constant parameters. It has been shown that the ABSURDIST iterative scheme is a close relative of the classic steepest descent method. It is, in fact, an optimization method in the space of  $n \times n$  correspondence matrices. It maximizes an “energy functional”: a quadratic function of the correspondence matrix that contains positive reward terms for a match between the structures of the two concept systems being aligned (the internal similarity—excitation) and for a match between the system alignment and the *a priori* external similarity matrix, as well as a penalty term for non-orthogonality of rows or columns of the correspondence matrix. The functional is maximized on the set of matrices whose elements are within the specified range, i.e. on a cube  $Q$  in  $R^{n^2}$ .

We know that if this functional was to be considered on the set of all permutation matrices, it achieves its maximum on that set on the permutation matrix (or matrices) that minimize relation mismatch between the graphs. Although we lack a strict proof, we conjecture that when  $\chi$  is high enough, this matrix will deliver the global maximum on  $Q$  as well.

The algorithm differs from the standard steepest descent in that the components of the update vector at each step are “damped” in such a way to ensure that the coefficients of the correspondence matrix stay in the prescribed range. As a result, the algorithm may converge either to the absolute maximum of the energy functional, if it is located within cube  $Q$ , or to a point on the border of the cube where a local maximum on the cube is reached.

Maximizing a quadratic polynomial on a cube in a linear space is the standard quadratic programming problem. It is well known that the local maxima may be numerous. Therefore, for a given set of parameters, the ABSURDIST algorithm may converge to various correspondence matrices (local convergence points); which one a particular iterative process converges to depends on the initial starting point.

We have made a study of ABSURDIST's convergence to several types of such local maxima located in various “corners” of the cube. In particular, we have shown that for important classes of concept systems, if the ratio of ABSURDIST parameters  $\chi/\beta$  is low enough, the algorithm will locally converge a matrix of all ones, or a matrix with a large block of all ones, will be a local convergence point. On the other hand, if  $\chi$  is high enough, some or even all permutation matrices will be local convergence points.

We have shown that in at least one practically important situation—trying to match, without the help of external similarity, two sparse isomorphous systems that have certain simple kinds of internal symmetry—any  $\chi/\beta$  setting that is high enough to ensure local convergence to the “correct” permutation matrix will also make all other  $2^n - 1$  permutation matrices local convergence points.

The presence of the external similarity may affect convergence, changing the convergence-point status of some “corners” of the cube  $Q$ , and altering the size of the “catchment area” of others. The amount of external similarity necessary to ensure convergence depends on the structure of the systems being matched. While sometimes a limited amount of external similarity (symmetry breaking) may be quite helpful, we suspect that in most cases the amount of external symmetry that needs to be provided to ensure convergence to the desired point may be quite significant.

The location of the convergence points of the ABSURDIST algorithm is not influenced by the choice of the learning rate, as long as it is guaranteed to converge in the first place; however, its speed is. We have provided guidelines for choosing the highest value of learning rate that still ensures convergence.

Nonetheless, the main problem with the algorithm remains: when started with a particular starting point, it would only find one local maximum of the energy functional; since in general there may be as many as  $2^n$ , or perhaps more, local maxima, checking them all may be exponentially complex. This is not entirely surprising, considering that both the general nonconvex quadratic programming problem and the problem of inexact graph matching are, in general, NP-hard.

In practice, the chances of the optimization algorithm such as ABSURDIST converging to the global maximum (or a similar value) when starting with a random point depend on the size of the “catchment area” of the global maximum. It may be interesting to consider how changing the  $\chi/\beta$  ratio during iterations (i.e., changing the functional we are optimizing) may improve our chances of finding the best solution, or a good one.

In another direction, ABSURDIST can be compared to an algorithm that maximizes a similar functional on a sphere (keeping the vector 2-norm of  $C$  within a limit, rather than maintaining individual bounds for each component). The latter, in the absence of the external similarity would simply converge to the eigenvector of  $\mathcal{A}$  corresponding to its largest eigenvalue. (Compare with similarity flooding in [3]). However, since such a vector may have negative components, it is not quite clear how to interpret it for the purposes of establishing a match between the systems.

## References

- [1] Ying Feng, Robert L. Goldstone, and Vladimir Menkov. Absurdist ii: A graph matching algorithm and its application to conceptual system translation. In *FLAIRS '04*, 2004.
- [2] Robert L. Goldstone and Brian J. Rogosky. Using relations within conceptual systems to translate across conceptual systems. *Cognition*, pages 295–320, 2002.
- [3] Sergey Melnik, Hector Garcia-Molina, and Erhard Rahm. Similarity flooding: a versatile graph matching algorithm and its application to schema matching. In *18th International Conference on Data Engineering (ICDE)*, pages 117–128, 2002.
- [4] Bruno T. Messmer. *Efficient Graph Matching Algorithms*. PhD thesis, Universität Bern, 1995.