Supplementary Material

A Closed Form Solutions for Maximum Violation Loss for Simple Implications

Let \( \phi_r(h_1, h_2) \) be a scoring function for a relation \( r \) defined over pairs of entity vectors \( h_1 \) and \( h_2 \), such as the scoring function in DISTMULT or COMPLEX. In the following, we assume that all entity embeddings live on a subspace \( \mathcal{U} \subseteq \mathbb{R}^k \). The subspace \( \mathcal{U} \) can either correspond to the unit sphere – i.e. \( \mathcal{U} \triangleq \{ h \mid \|h\|_2 = 1 \} \) – or to the unit cube – i.e. \( \mathcal{U} \triangleq \{ h \mid h \in [0,1]^k \} \).

Let us consider a mapping \( \mathcal{S} : \mathcal{V} \mapsto \mathbb{R}^k \) from variables to \( k \)-dimensional embeddings, where \( h_i = \mathcal{S}(X_i), \forall i \).

Given a clause expressing a simple implication in the form \( \mathcal{U} \subseteq \text{IST} \) function in \( \phi \), let \( \delta \) where \( \delta \) is the corresponding derivations for different choices of the entity embedding subspace \( \mathcal{U} \).

A.1 DISTMULT

In the following, we focus on the Bilateral-Diagonal model (DISTMULT), proposed by [Yang et al. 2015], and provide the corresponding derivations for different choices of the entity embedding subspace \( \mathcal{U} \).

We want to solve the following optimisation problem:

\[
\mathcal{J}^\text{max} = \max_{h_1, h_2 \in \mathcal{U}} \left( \phi_b(h_1, h_2) - \phi_r(h_1, h_2) \right)
\]

where \( \delta \triangleq \theta_b - \theta_r \).

A.1.1 Unit Sphere

Assume that the subspace \( \mathcal{U} \) corresponds to the unit sphere, i.e. \( \forall x \in \mathcal{E} : \|x\|_2^2 = 1 \). The Lagrangian is:

\[
\mathcal{L} = -\langle \delta, h_1, h_2 \rangle + \lambda_1 (\|h_1\|_2^2 - 1) + \lambda_2 (\|h_2\|_2^2 - 1).
\]

Imposing stationarity: \( \nabla_{h_1} \mathcal{L} = 0 \) and \( \nabla_{h_2} \mathcal{L} = 0 \) gives:

\[
-\delta \odot h_2 + 2\lambda_1 h_1 = 0
\]

\[
-\delta \odot h_1 + 2\lambda_2 h_2 = 0
\]

in which \( \odot \) denotes the component-wise multiplication. For \( \lambda_1 \neq 0 \), a simple substitution leads to:

\[
-\delta^2 \odot h_2 + 4\lambda_1 \lambda_2 h_2 = 0,
\]

with the notation \( \delta \odot \delta = \delta^2 \). As a result, \( 4\lambda_1 \lambda_2 = \delta^2_i \) for components \( i \) with \( h_{1,i} \neq 0 \). Given the symmetry of the equations, the same requirements \( 4\lambda_1 \lambda_2 = \delta^2_i \) hold for components \( h_{2,i} \neq 0 \).

We search for \( h_1 \) and \( h_2 \), such that \( \delta^2_i \) is constant for their non-zero components. Construct \( h_1 \) and \( h_2 \) such that only their component \( j \) is non-zero: \( \forall i \neq j : h_{1,i} = h_{2,i} = 0 \), whereas \( h_{1,j} = \pm 1, h_{2,j} = \pm 1 \) (given the unit sphere constraint). The contribution of component \( j \) to \( \langle \delta, h_1, h_2 \rangle \) depends on \( h_{1,i}, h_{2,i} \) which can take values \( \pm 1 \). As a result:

\[
\mathcal{J}^\text{max} = \max_j |\delta_j|.
\]
In the case where several components \( \delta_j \) have the same value, then all \( h_{1,i} \) and \( h_{2,i} \) need to be zero for \( i \neq j \), but due to the normalisation constraint, the highest value of \( \sum_j h_{1,j} h_{2,j} \) is found for a single index \( j \) if both components take the value \( \pm 1 \).

Finally, since \( J^{\text{max}} \) is always non-negative, we find:

\[
J^{\text{max}}_I = \max_j (\theta_{h,j} - \theta_{r,j})
\]

### A.1.2 Unit cube

For the entity embeddings to be constrained in the unit cube, their subspace is set to \( \mathcal{U} = [0, 1]^k \subset \mathbb{R}^k \). This corresponds to reducing the entity embeddings to approximately Boolean embeddings (see [Demeester et al., 2016]).

The Lagrangian becomes:

\[
L = -\langle \delta, h_1, h_2 \rangle + \sum_i [\mu_1 \odot (h_1 - 1) + \mu_2 \odot (h_2 - 1)]_i
\]

with \( \forall i : h_{1,i} - 1 \leq 0, h_{2,i} - 1 \leq 0 \) (primal feasibility), \( \forall i : \mu_{1,i} \geq 0, \mu_{2,i} \geq 0 \) (dual feasibility), \( \forall i : \mu_{1,i}(h_{1,i} - 1) = 0, \mu_{2,i}(h_{2,i} - 1) = 0 \) (complementary slackness). In fact, we should add KKT multipliers for the conditions \(-h_{1,i} \leq 0, -h_{2,i} \leq 0\) as well. These don’t change the results, if we ensure the unit cube restrictions are satisfied.

Imposing stationarity, or \( \nabla_h L = 0 \) and \( \nabla_{h_2} L = 0 \), we get:

\[
-\delta \odot h_2 + \mu_1 = 0 \\
-\delta \odot h_1 + \mu_2 = 0
\]

This can be solved component-wise, or for any component \( i \):

\[
\mu_{1,i} = \delta_i h_{2,i} \\
\mu_{2,i} = \delta_i h_{1,i}
\]

Dual feasibility dictates that \( \mu_{1,i} \geq 0 \) and \( \mu_{2,i} \geq 0 \), such that \( h_{2,i} = 0 \) and \( h_{1,i} = 0 \) for any components where \( \delta_i < 0 \).

For the other components, we have \( \delta_i \geq 0 \), such that while satisfying the unit cube constraints the highest value of the objective becomes \( J^{\text{max}}_I = \delta_i \) for \( h_{1,i} = h_{2,i} = 1 \).

Finally, we find:

\[
J^{\text{max}}_I = \sum_j \max(0, \theta_{b,j} - \theta_{r,j}),
\]

which is exactly the same expression as the lifted loss for simple implications with unit cube entity embeddings introduced by [Demeester et al., 2016] for MODEL F (which can be seen as a special case of DISTMULT where the subject embeddings are replaced by entity pair embeddings, and all object embedding components are set to 1).

### A.2 COMPLEX

In COMPLEX, proposed by [Trouillon et al., 2016], the scoring function \( \phi \) is defined as follows:

\[
\phi^\theta(h_1, h_2) \triangleq (\theta_r, h_1, \overline{h_2})^R
\]

where \( x^R \triangleq \text{Re}(x) \) and \( x^I \triangleq \text{Im}(x) \) denote the real and imaginary part of \( x \), respectively. In the following, we analyse the cases where entity embeddings live on the unit sphere (i.e. \( \forall x \in \mathcal{E} : ||h_x||_2 = 1 \)) and in the unit cube (i.e. \( \forall x \in \mathcal{E} : h_x^R \in [0, 1]^k, h_x^I \in [0, 1]^k \)).

We want to solve the following maximisation problem:

\[
J = \max_{h_1, h_2 \in \mathcal{U}} (\phi(h_1, h_2) - \phi_r(h_1, h_2))
\]

\[
= \max_{h_1, h_2 \in \mathcal{U}} (\delta, h_1, \overline{h_2})^R,
\]
with the complex vector $\delta = \theta_b - \theta_s$.

To keep notations simple, we will continue to work with complex vectors as much as possible. For the Lagrangian $\mathcal{L}$, the pair of stationarity equations with respect to the real and imaginary part of a variable (say $h_1$), can be taken together as follows:

$$\nabla_{h_1} \mathcal{L} + j \nabla_{h_2} \mathcal{L} = 0$$

and we introduce the notation $\nabla_{h_i} \mathcal{L} \equiv \nabla_{h_1} \mathcal{L} + j \nabla_{h_2} \mathcal{L}$, with $j$ the imaginary unit. For the remainder, we need the following:

$$\nabla_{h_1} (\delta, h_1, \overline{h_2})^R = \frac{1}{2} \nabla_{h_1} \left( (\delta, h_1, \overline{h_2}) + (\overline{\delta}, h_1, h_2) \right)$$

$$= \frac{1}{2} (\delta \circ h_2 + j (j\delta \circ \overline{h_2}) + \overline{\delta} \circ h_2 + j (-j\overline{\delta} \circ h_2))$$

$$= \overline{\delta} \circ h_2$$

$$\nabla_{h_2} (\delta, h_1, \overline{h_2})^R = \delta \circ h_1$$

A.2.1 Unit sphere

We now restrict the subject and object embeddings to live on the unit sphere, i.e. $\forall x \in \mathcal{E} : \|h_x\|^2 = 1$. Given that the adversarial entity embeddings need to live on the unit sphere, the Lagrangian can be defined as follows:

$$\mathcal{L}(h_1, h_2, \lambda_1, \lambda_2) = - (\delta, h_1, \overline{h_2})^R$$

$$+ \lambda_1 \left( \|h_1\|^2 - 1 \right)$$

$$+ \lambda_2 \left( \|h_2\|^2 - 1 \right)$$

with real-valued Lagrange multipliers $\lambda_1$ and $\lambda_2$, and in which $\|h_1\|^2 \triangleq \|h_1^R\|^2 + \|h_1^I\|^2$ is the L2 norm of the complex vector $h_1$. With the expressions for $\nabla_{h_1} (\delta, h_1, \overline{h_2})^R$ and $\nabla_{h_2} (\delta, h_1, \overline{h_2})^R$ above, the stationarity conditions can be written out as follows:

$$-\overline{\delta} \circ h_2 + 2\lambda_1 h_1 = 0$$

$$-\delta \circ h_1 + 2\lambda_2 h_2 = 0$$

Substituting into each other, we find:

$$4\lambda_1 \lambda_2 h_1 = \overline{\delta} \circ \delta \circ h_1, \quad (\lambda_2 \neq 0)$$

$$4\lambda_1 \lambda_2 h_2 = \delta \circ \overline{\delta} \circ h_2, \quad (\lambda_1 \neq 0)$$

As a result, we require $4\lambda_1 \lambda_2 = |\delta_i|^2$ for components $i$ with $h_{1,i} \neq 0$ or $h_{2,i} \neq 0$. As in the case with DistMult, take $h_{1,i} = h_{2,i} = 0$ for each component $i \neq j$, such that for component $j$, we need $|h_{1,j}| = |h_{2,j}| = 1$, such that $|h_{1,j} h_{2,j}| = 1$. In order to maximise the contribution of that component to the loss, we choose the argument of the complex number $h_{1,j} h_{2,j}$ such that $\delta_j |h_{1,j} h_{2,j}|$ falls on the positive real axis. As a result:

$$J^\text{max}_2 = \max_j |b_j - r_j| = \max_j \sqrt{(\theta^R_{b,j} - \theta^R_{r,j})^2 + (\theta^I_{b,j} - \theta^I_{r,j})^2}$$

A.2.2 Unit cube

This case can be solved with the KKT conditions again, but instead we provide a shorter, less formal, derivation. It is clear that we can maximise the objective by maximising each component independently. For component $i$ we need to optimise the following:

$$\delta^R_{1,h_{1,i}} h_{2,i}^R + \delta^I_{1,h_{1,i}} h_{2,i}^I + \delta^R_{1,h_{1,i}} h_{2,i}^R - \delta^I_{1,h_{1,i}} h_{2,i}^I$$

Regrouping gives:

$$\alpha \delta^R_i + \beta \delta^I_i,$$
with $\alpha = h_{1,i}^R h_{2,i}^R + h_{1,i}^I h_{2,i}^I$ and $\beta = h_{1,i}^R h_{2,i}^I - h_{1,i}^I h_{2,i}^R$. We know $0 \leq \alpha \leq 2$, $-1 \leq \beta \leq 1$ and $\alpha + |\beta| \leq 2$.

This allows maximising the objective as follows:

$$J^\text{max} = \sum_i \max(\delta_i^R, 0) + \max(\delta_i^R, |\delta_i^I|)$$

### B Simple Implications with Swapped Arguments

Given a clause expressing a simple implication with swapped arguments, in the form $b(X_1, X_2) \Rightarrow r(X_2, X_1)$, we would like to maximise the inconsistency loss $J_I$ associated to the clause, i.e.:

$$J^I = \max \left(0, J^\text{max} \right)$$

with $J^\text{max} = \max_{h_i, h_2 \in U} \left( \phi_b(h_1, h_2) - \phi_r(h_2, h_1) \right)$.

#### B.1 DISTMULT

Due to symmetry, the same close form expressions as for the simple implications hold.

#### B.2 COMPLEX

We want to solve the following maximisation problem:

$$J^\text{max} = \max_{h_1, h_2 \in U} \left( \phi_b(h_1, h_2) - \phi_r(h_2, h_1) \right)$$

$$= \max_{h_1, h_2 \in U} \left( \theta_b, h_1, h_2 \right)^R - \left( \theta_r, h_2, h_1 \right)^R$$

$$= \max_{h_1, h_2 \in U} \left( \theta_b - \theta_r, h_1, h_2 \right)^R$$

$$= \max_{h_1, h_2 \in U} \left( \zeta, h_1, h_2 \right)^R$$

This has the same form as the simple implications case, but with $\theta_r$ replaced by $\theta_r^R$, or, more specifically, $\theta_r^I$ by $-\theta_r^I$.

#### B.2.1 Unit sphere

Under unit sphere constraints, $J^\text{max}$ has the following value:

$$J^\text{max} = \max_i \sqrt{(\theta_{b,i}^R - \theta_{r,i}^R)^2 + (\theta_{b,i}^I + \theta_{r,i}^I)^2}$$

#### B.2.2 Unit cube

With $\zeta_i^R = \theta_{b,i}^R - \theta_{r,i}^R$ and $\zeta_i^I = \theta_{b,i}^I + \theta_{r,i}^I$:

$$J^\text{max} = \sum_i \max(\zeta_i^R, 0) + \max(\zeta_i^R, |\zeta_i^I|)$$

### C Symmetry

Given a clause expressing a simple implication with swapped arguments, in the form $r(X_1, X_2) \Rightarrow r(X_2, X_1)$, we would like to maximise the inconsistency loss $J_I$ associated to the clause, i.e.:

$$J^I = \max \left(0, J^\text{max} \right)$$

with $J^\text{max} = \max_{h_1, h_2 \in U} \left( \phi_r(h_1, h_2) - \phi_r(h_2, h_1) \right)$. 

Note that this is a special case of Appendix B where \( r = b \).

### C.1 DistMult

Since \textsc{DistMult} is symmetric, the gradient for symmetry clauses is zero, \textit{i.e.}, all relations already satisfy symmetry.

### C.2 Complex

We want to solve the following maximisation problem:

\[
\mathcal{J}^\text{max} = \max_{h_1, h_2 \in U} \left( \phi_r(h_1, h_2) - \phi_r(h_2, h_1) \right)
\]

\[
= \max_{h_1, h_2 \in U} \langle \theta - \theta^T, h_1 \rangle \langle \theta - \theta^T, h_2 \rangle^R
\]

**C.2.1 Unit sphere**

From Eq. (4) with \( \theta^R_{r,i} = \theta^R_{b,i} \) and \( \theta^I_{r,i} = \theta^I_{b,i} \) we get:

\[
\mathcal{J}^\text{max} = \max_i 2|\theta^I_{r,i}|
\]

**C.2.2 Unit cube**

Similarly, with \( \theta^R_{r,i} = \theta^R_{b,i} \) and \( \theta^I_{r,i} = \theta^I_{b,i} \) we get:

\[
\mathcal{J}^\text{max} = 2 \sum_i |\theta^I_{r,i}|
\]

### D Link Prediction Results

Table 8: Link prediction results on the Test-I, Test-II and and Test-ALL on FB122, filtered setting.

<table>
<thead>
<tr>
<th>Model</th>
<th>FB122</th>
<th></th>
<th></th>
<th>WMN18</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Test-I</td>
<td>Test-II</td>
<td>Test-ALL</td>
<td>MRR</td>
<td>Test-I</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Hits@N (%)</td>
<td>Hits@N (%)</td>
<td>Hits@N (%)</td>
<td>MRR</td>
<td>Hits@N (%)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>3 5 10</td>
<td>3 5 10</td>
<td>3 5 10</td>
<td></td>
<td>3 5 10</td>
</tr>
</tbody>
</table>

**Supplementary References**