The Closest String problem (CSP) is a combinatorial optimization problem that receives as input a set of strings of the same length and seeks a string whose distance from the given strings is minimal. The distance from a solution to a given string is measured by the Hamming distance. The distance from the solution to the farthest input string is considered as the objective value of the solution, which is to be minimized. The CSP is NP-hard [Frances and Litman, 97].

NP-hard problems hierarchy for the CSP [Lanctot et al., 2003]

\[
\text{3SAT} \rightarrow \text{DSSP} \rightarrow \text{FSP} \rightarrow \text{FSubSP} \rightarrow \text{CSP} \rightarrow \text{CSubSP} \rightarrow \text{DSSP}
\]

**Theorem [O.L.Vilca]**

Let a CSP instance with 3-sequences, which denotes 3-CSP, \( S = \{s^i \in \Sigma^m, 1 \leq i \leq 3\} \) with alphabet \( |\Sigma| > 2 \), so the 3-CSP-A algorithm always finds an exact solution to 3-CSP [Liu et al., 2001] for Binary case.

**Proof**

This proof is by construction method, based on normalized instance [Gramm et al., 2001],

\[
\frac{\sum_{i \in \{1,2,3\}} \sum_{j \in \{1,2,3\}} x_{ij} \in \{0,1\}}{\text{Number of columns}}
\]

\[
\phi(c_i) = \begin{cases} 
\alpha & \text{if } c_i \in C_1 : \max(c_i, c_i') = c_i \\
\beta & \text{else if } c_i \in C_2 : \max(c_i, c_i') = c_i' \\
\gamma & \text{else if } c_i \in C_3 : \max(c_i, c_i') = \{c_1, c_2\} \\
\end{cases}
\]

After that, we reduce into five different cases

\[
\begin{align*}
\text{Case 1: } & \quad \alpha \\
\text{Case 2: } & \quad \beta \\
\text{Case 3: } & \quad \gamma \\
\text{Case 4: } & \quad \alpha \beta \\
\text{Case 5: } & \quad \beta \alpha \\
\end{align*}
\]

For blocks of 2 and 3-length, with \( \{i, j, k\} \in \{2, 3, 4\} \), we have

\[
\begin{align*}
\sum_{i=1}^{2} \sum_{j=1}^{2} x_{ij} & = \sum_{i=1}^{3} \sum_{j=1}^{3} x_{ij} \\
\sum_{i=1}^{2} \sum_{j=1}^{2} y_{ij} & = \sum_{i=1}^{3} \sum_{j=1}^{3} y_{ij}
\end{align*}
\]

So in truth we are interested in the cases when the Hamming distance is equal to 1. Let \( t_i \) the number of times that \( v_i \) is repeated, \( b_i \leq t_i \leq l_i \), then \( l_i = l_i - b_i \). Let \( \rho_{\max} \) be \( \frac{l_i m}{t_i} \) for \( 2 \leq t_i \leq 4 \).

\[
\begin{align*}
\text{Case 1: } & \quad \alpha \\
\text{Case 2: } & \quad \beta \\
\text{Case 3: } & \quad \gamma \\
\text{Case 4: } & \quad \alpha \beta \\
\text{Case 5: } & \quad \beta \alpha
\end{align*}
\]

Let \( r \) a string that represents an optimal solution of 3-CSP-A, with \( 1 \leq i \leq m \), we have:

\[
\begin{align*}
s^1_i = s^2_i & \quad t_i = s^3_i \\
s^1_i = s^2_i \neq s^3_i & \quad \rho_{\max} > 0 \quad t_i = s^3_i \\
s^1_i \neq s^2_i = s^3_i & \quad \rho_{\max} = 0 \quad t_i = s^2_i \\
s^1_i \neq s^2_i \neq s^3_i & \quad \rho_{\max} > 0 \quad t_i = s^3_i \\
s^1_i \neq s^2_i \neq s^3_i & \quad \rho_{\max} > 0 \quad t_i = s^3_i \\
s^1_i \neq s^2_i \neq s^3_i & \quad \rho_{\max} > 0 \quad t_i = s^3_i \\
s^1_i \neq s^2_i \neq s^3_i & \quad \rho_{\max} > 0 \quad t_i = s^3_i
\end{align*}
\]

Thus the theorem holds.

**Computational results**

Linear time algorithm 3-CSP-A, for 3-sequences.

**Concluding remark**

We proposed an exact algorithm for the special case of CSP with 3-sequences and alphabet size \(|\Sigma| > 2\) and gave the corresponding theoretical analysis.

**Others names for the CSP.**

Year | Name | References
--- | --- | ---
1997 | Minimum Radius Problem | [Frances and Litman, 97]
1999 | Hamming Center Problem | [Gasieniec et al., 99]
2001 | Center String Problem | [Gramm et al., 2001]
2003 | Consensus String Problem | [Lanctot et al., 2003]
2008 | Motif Finding Problem | [Gomes et al., 2008]

**References**