Parameterized Algorithms for Matrix Completion with Radius Constraints

Tomohiro Koana[1] Vincent Froese[1]
Rolf Niedermeier[1]

[1]Techqnische Universität Berlin, Faculty IV, Algorithmics and Computational Complexity

May 12, 2020

Problem Statement

For $i = 1 \dots n$:

$S[i]$ string on alphabet Σ admitting wildcards $*$

$$S = \begin{pmatrix} \mathtt{b} & \mathtt{a} & \mathtt{n} & \mathtt{a} & \mathtt{n} & \mathtt{a} \\ \mathtt{a} & * & \mathtt{n} & * & \mathtt{n} & * \\ \mathtt{b} & * & * & \mathtt{c} & \mathtt{z} & \mathtt{a} \\ \mathtt{z} & \mathtt{a} & \mathtt{z} & * & \mathtt{n} & * \\ \mathtt{b} & \mathtt{a} & * & \mathtt{a} & \mathtt{c} & \mathtt{a} \\ \mathtt{b} & * & \mathtt{n} & \mathtt{c} & * & \mathtt{z} \end{pmatrix} \in \left(\Sigma \cup \{*\}\right)^{n \times \ell}$$

Problem Statement

For $i = 1 \dots n$:

$S[i]$ string on alphabet Σ admitting wildcards $*$

$$S = \begin{pmatrix} \mathtt{b} & \mathtt{a} & \mathtt{n} & \mathtt{a} & \mathtt{n} & \mathtt{a} \\ \mathtt{a} & * & \mathtt{n} & * & \mathtt{n} & * \\ \mathtt{b} & * & * & \mathtt{c} & \mathtt{z} & \mathtt{a} \\ \mathtt{z} & \mathtt{a} & \mathtt{z} & * & \mathtt{n} & * \\ \mathtt{b} & \mathtt{a} & * & \mathtt{a} & \mathtt{c} & \mathtt{a} \\ \mathtt{b} & * & \mathtt{n} & \mathtt{c} & * & \mathtt{z} \end{pmatrix} \in (\Sigma \cup \{*\})^{n \times \ell}$$

Minimum Radius Matrix Completion (MinRMC)

Input: Matrix $S \in (\Sigma \cup \{*\})^{n \times \ell}$ and $d \in \mathbb{N}$.
Output: $\exists T$ completion of S and $\exists v \in \Sigma^{\ell}$ s.t. $\delta(v, T) \leq d$?

Problem Statement

For $i = 1 \ldots n$:

$S[i]$ string on alphabet Σ admitting wildcards $*$

$$S = \begin{pmatrix} \mathtt{b} & \mathtt{a} & \mathtt{n} & \mathtt{a} & \mathtt{n} & \mathtt{a} \\ \mathtt{a} & * & \mathtt{n} & * & \mathtt{n} & * \\ \mathtt{b} & * & * & \mathtt{c} & \mathtt{z} & \mathtt{a} \\ \mathtt{z} & \mathtt{a} & \mathtt{z} & * & \mathtt{n} & * \\ \mathtt{b} & \mathtt{a} & * & \mathtt{a} & \mathtt{c} & \mathtt{a} \\ \mathtt{b} & * & \mathtt{n} & \mathtt{c} & * & \mathtt{z} \end{pmatrix} \in \left(\Sigma \cup \{*\}\right)^{n \times \ell}$$

Constraint Radius Matrix Completion (ConRMC)

Input: Matrix $S \in \left(\Sigma \cup \{*\}\right)^{n \times \ell}$ and $d_1 \ldots d_n \in \mathbb{N}$.
Output: $\exists T$ completion of S and $\exists v \in \Sigma^{\ell}$ s.t. $\forall i \; \delta\left(v, T[i]\right) \leq d_i$?

Related Work

Stringology and ML meet in the middle!

Related Work

Stringology and ML meet in the middle!

ICML '18: Parameterized Algorithms for Matrix Completion

Dataset contains missing entries

Fill the gaps optimizing some metrics (e.g. rank of the matrix)

Related Work

Stringology and ML meet in the middle!

ICML '18: Parameterized Algorithms for Matrix Completion

Dataset contains missing entries

Fill the gaps optimizing some metrics (e.g. rank of the matrix)

CPM '14: Parameterized Closest String with Wildcards

Closest string $\equiv$ MinRMC for a wildcard-free S

Applications in Bioinformatics

- Motif finding
- PCR primer design
- Genetic probe design

Hardness Results

All Problems are Intractable

- Closest String is $\mathcal{NP}$-hard, even for $|\Sigma| = 2$
- Closest String $\preceq$ MinRMC $\preceq$ ConRMC

Hardness Results

All Problems are Intractable

- Closest String is $\mathcal{NP}$-hard, even for $|\Sigma| = 2$
- Closest String $\preceq$ MinRMC $\preceq$ ConRMC

Fixed Parameter Tractability

$f(n, k) = \mathcal{O}\left(g(k) n^{\mathcal{O}(1)}\right)$

Hardness Results

All Problems are Intractable

- Closest String is $\mathcal{NP}$-hard, even for $|\Sigma| = 2$
- Closest String $\preceq$ MinRMC $\preceq$ ConRMC

Fixed Parameter Tractability

$$f(n,k) = \mathcal{O}\left(g(k) n^{\mathcal{O}(1)}\right) \iff f(n,k) = \mathcal{O}^*\left(g(k)\right)$$

State of Art

CPM '14 solved MinRMC in:

- $\mathcal{O}^*\left(2^{\ell^2/2}\right)$
- $\mathcal{O}\left(n\ell^2\right)$ for $d = 1$
- $\mathcal{O}^*\left(|\Sigma|^k \cdot d^d\right)$

Here we solve MinRMC in:

- $\mathcal{O}^*\left(\ell^\ell\right)$
- $\mathcal{O}\left(n\ell\right)$ for $d = 1$
- $\mathcal{O}^*\left(|\Sigma|^{d+k} \cdot 2^{4k+d}\right)$

State of Art

CPM '14 solved MinRMC in:

- $\mathcal{O}^*\left(2^{\ell^2/2}\right)$
- $\mathcal{O}\left(n\ell^2\right)$ for $d = 1$
- $\mathcal{O}^*\left(|\Sigma|^k \cdot d^d\right)$

Here we solve MinRMC in:

- $\mathcal{O}^*\left(\ell^\ell\right)$
- $\mathcal{O}\left(n\ell\right)$ for $d = 1$
- $\mathcal{O}^*\left(|\Sigma|^{d+k} \cdot 2^{4k+d}\right)$

A bug in CPM '14 $\mathcal{O}^*\left((d+1)^{d+k}\right)$ algorithm is fixed

All results are proven for the more general **ConRMC**

Parameter: Number ℓ of Columns

$$S = \begin{pmatrix} \mathtt{b} & \mathtt{a} & \mathtt{n} & \mathtt{a} & \mathtt{n} & \mathtt{a} \\ \mathtt{a} & * & \mathtt{n} & * & \mathtt{n} & * \\ \mathtt{b} & * & * & \mathtt{c} & \mathtt{z} & \mathtt{a} \\ \mathtt{z} & \mathtt{a} & \mathtt{z} & * & \mathtt{n} & * \\ \mathtt{b} & \mathtt{a} & * & \mathtt{a} & \mathtt{c} & \mathtt{a} \\ \mathtt{b} & * & \mathtt{n} & \mathtt{c} & * & \mathtt{z} \end{pmatrix} \quad d = \begin{pmatrix} 4 \\ 3 \\ 4 \\ 2 \\ 5 \\ 3 \end{pmatrix}$$

ConRMC: find $v \in \Sigma^\ell$ s.t. $\exists T$ completion of S, $\forall i \; \delta(v, T[i]) \leq d_i$

Define $P_*(S[i]) = \{j \leq \ell \mid S[i,j] = *\}$, $P_*(S[2]) = \{2, 4, 6\}$

Parameter: Number ℓ of Columns

$$S = \begin{pmatrix} \texttt{b} & \texttt{a} & \texttt{n} & \texttt{a} & \texttt{n} & \texttt{a} \\ \texttt{a} & * & \texttt{n} & * & \texttt{n} & * \\ \texttt{b} & * & * & \texttt{c} & \texttt{z} & \texttt{a} \\ \texttt{z} & \texttt{a} & \texttt{z} & * & \texttt{n} & * \\ \texttt{b} & \texttt{a} & * & \texttt{a} & \texttt{c} & \texttt{a} \\ \texttt{b} & * & \texttt{n} & \texttt{c} & * & \texttt{z} \end{pmatrix} \quad d = \begin{pmatrix} 4 \\ 3 \\ 4 \\ 2 \\ 5 \\ 3 \end{pmatrix}$$

ConRMC: find $v \in \Sigma^\ell$ s.t. $\exists T$ completion of S, $\forall i\ \delta(v, T[i]) \leq d_i$

Define $P_*(S[i]) = \{j \leq \ell \mid S[i,j] = *\}$, $P_*(S[2]) = \{2, 4, 6\}$

Parameter: Number ℓ of Columns

$$S = \begin{pmatrix} \texttt{b} & \texttt{a} & \texttt{n} & \texttt{a} & \texttt{n} & \texttt{a} \\ \texttt{a} & * & \texttt{n} & * & \texttt{n} & * \\ \texttt{b} & * & * & \texttt{c} & \texttt{z} & \texttt{a} \\ \texttt{z} & \texttt{a} & \texttt{z} & * & \texttt{n} & * \\ \texttt{b} & \texttt{a} & * & \texttt{a} & \texttt{c} & \texttt{a} \\ \texttt{b} & * & \texttt{n} & \texttt{c} & * & \texttt{z} \end{pmatrix} \quad d = \begin{pmatrix} 4 \\ 3 \\ 4 \\ 2 \\ 4 \\ 3 \end{pmatrix}$$

Any $v \in \Sigma^\ell$ is a center $\iff \forall i \leq n \quad \ell - P_*(S[i]) \leq d_i$

If holds:

return **True**

Else:

$\exists R_i \subseteq \{1 \ldots \ell\} \setminus P_*(S[i])$ s.t. $|R_i| = d_i + 1$

Parameter: Number ℓ of Columns

$$S = \begin{pmatrix} \mathtt{b} & \mathtt{a} & \mathtt{n} & \mathtt{a} & \mathtt{n} & \mathtt{a} \\ \mathtt{a} & * & \mathtt{n} & * & \mathtt{n} & * \\ \mathtt{b} & * & * & \mathtt{c} & \mathtt{z} & \mathtt{a} \\ \mathtt{z} & \mathtt{a} & \mathtt{z} & * & \mathtt{n} & * \\ \color{red}{\mathtt{b}} & \color{red}{\mathtt{a}} & \color{blue}{*} & \color{red}{\mathtt{a}} & \color{red}{\mathtt{c}} & \color{red}{\mathtt{a}} \\ \mathtt{b} & * & \mathtt{n} & \mathtt{c} & * & \mathtt{z} \end{pmatrix} \quad d = \begin{pmatrix} 4 \\ 3 \\ 4 \\ 2 \\ \color{green}{4} \\ 3 \end{pmatrix}$$

$$\text{Any } v \in \Sigma^{\ell} \text{ is a center } \iff \forall i \leq n \quad \ell - P_*(S[i]) \leq d_i$$

If holds:

return **True**

Else:

$\exists \color{red}{R_i} \subseteq \{1 \ldots \ell\} \setminus \color{blue}{P_*(S[i])}$ s.t. $|\color{red}{R_i}| = \color{green}{d_i} + 1$

Parameter: Number ℓ of Columns

$\exists R_i \subseteq \{1 \dots \ell\} \setminus P_* (S[i])$ s.t. $|R_i| = d_i + 1$

$$S = \begin{pmatrix} \mathtt{b} & \mathtt{a} & \mathtt{n} & \mathtt{a} & \mathtt{n} & \mathtt{a} \\ \mathtt{a} & * & \mathtt{n} & * & \mathtt{n} & * \\ \mathtt{b} & * & * & \mathtt{c} & \mathtt{z} & \mathtt{a} \\ \mathtt{z} & \mathtt{a} & \mathtt{z} & * & \mathtt{n} & * \\ \mathtt{b} & \mathtt{a} & * & \mathtt{a} & \mathtt{c} & \mathtt{a} \\ \mathtt{b} & * & \mathtt{n} & \mathtt{c} & * & \mathtt{z} \end{pmatrix} \quad d = \begin{pmatrix} 4 \\ 3 \\ 4 \\ 2 \\ 4 \\ 3 \end{pmatrix}$$

$v \in \Sigma^\ell$ is a center $\implies \exists j \in R_i$ s.t. $v[j] = S[i, j]$

Parameter: Number ℓ of Columns

$\exists R_i \subseteq \{1 \ldots \ell\} \setminus P_* (S[i])$ s.t. $|R_i| = d_i + 1$

$$S = \begin{pmatrix} \mathtt{b} & \mathtt{a} & \mathtt{n} & \mathtt{a} & \mathtt{n} & \mathtt{a} \\ \mathtt{a} & * & \mathtt{n} & * & \mathtt{n} & * \\ \mathtt{b} & * & * & \mathtt{c} & \mathtt{z} & \mathtt{a} \\ \mathtt{z} & \mathtt{a} & \mathtt{z} & * & \mathtt{n} & * \\ \mathtt{b} & \mathtt{a} & * & \mathtt{a} & \mathtt{c} & \mathtt{a} \\ \mathtt{b} & * & \mathtt{n} & \mathtt{c} & * & \mathtt{z} \end{pmatrix} \quad d = \begin{pmatrix} 4 \\ 3 \\ 4 \\ 2 \\ 4 \\ 3 \end{pmatrix}$$

$$v \in \Sigma^\ell \text{ is a center } \implies \exists j \in R_i \text{ s.t. } v[j] = S[i,j]$$

For $j \in R_i$:
Set $v[j] = S[i,j]$, process j-th column and recurse

Parameter: Number ℓ of Columns

$\exists R_i \subseteq \{1 \dots \ell\} \setminus P_* (S[i])$ s.t. $|R_i| = d_i + 1$

$$S = \begin{pmatrix} \boxed{\mathtt{b}} & \mathtt{a} & \mathtt{n} & \mathtt{a} & \mathtt{n} & \mathtt{a} \\ \boxed{\mathtt{a}} & * & \mathtt{n} & * & \mathtt{n} & * \\ \boxed{\mathtt{b}} & * & * & \mathtt{c} & \mathtt{z} & \mathtt{a} \\ \boxed{\mathtt{z}} & \mathtt{a} & \mathtt{z} & * & \mathtt{n} & * \\ \boxed{\mathtt{b}} & \mathtt{a} & * & \mathtt{a} & \mathtt{c} & \mathtt{a} \\ \boxed{\mathtt{b}} & * & \mathtt{n} & \mathtt{c} & * & \mathtt{z} \end{pmatrix} \quad d = \begin{pmatrix} 4 \\ 3 \\ 4 \\ 2 \\ 4 \\ 3 \end{pmatrix}$$

$$v \in \Sigma^\ell \text{ is a center } \implies \exists j \in R_i \text{ s.t. } v[j] = S[i,j]$$

For $j \in R_i$:
Set $v[j] = S[i,j]$, process j-th column and recurse

Parameter: Number ℓ of Columns

$\exists R_i \subseteq \{1 \dots \ell\} \setminus P_*(S[i])$ s.t. $|R_i| = d_i + 1$

$$S = \begin{pmatrix} \texttt{b} & \texttt{a} & \texttt{n} & \texttt{a} & \texttt{n} & \texttt{a} \\ \texttt{a} & * & \texttt{n} & * & \texttt{n} & * \\ \texttt{b} & * & * & \texttt{c} & \texttt{z} & \texttt{a} \\ \texttt{z} & \texttt{a} & \texttt{z} & * & \texttt{n} & * \\ \texttt{b} & \texttt{a} & * & \texttt{a} & \texttt{c} & \texttt{a} \\ \texttt{b} & * & \texttt{n} & \texttt{c} & * & \texttt{z} \end{pmatrix} \quad d = \begin{pmatrix} 4 \\ 3 \\ 4 \\ 2 \\ 4 \\ 3 \end{pmatrix} \quad d' = \begin{pmatrix} 4 \\ 2 \\ 4 \\ 1 \\ 4 \\ 3 \end{pmatrix}$$

$$v \in \Sigma^\ell \text{ is a center } \implies \exists j \in R_i \text{ s.t. } v[j] = S[i,j]$$

For $j \in R_i$:
Set $v[j] = S[i,j]$, process j-th column and recurse

Parameter: Number ℓ of Columns

$$S = \begin{pmatrix} \texttt{b} & \texttt{a} & \texttt{n} & \texttt{a} & \texttt{n} & \texttt{a} \\ \texttt{a} & * & \texttt{n} & * & \texttt{n} & * \\ \texttt{b} & * & * & \texttt{c} & \texttt{z} & \texttt{a} \\ \texttt{z} & \texttt{a} & \texttt{z} & * & \texttt{n} & * \\ \texttt{b} & \texttt{a} & * & \texttt{a} & \texttt{c} & \texttt{a} \\ \texttt{b} & * & \texttt{n} & \texttt{c} & * & \texttt{z} \end{pmatrix} \quad d = \begin{pmatrix} 4 \\ 3 \\ 4 \\ 2 \\ 4 \\ 3 \end{pmatrix} \quad d' = \begin{pmatrix} 4 \\ 2 \\ 4 \\ 1 \\ 4 \\ 3 \end{pmatrix}$$

$$T(n, \ell) = (\widetilde{d} + 1)\, T(n, \ell - 1) + n\ell$$

Parameter: Number ℓ of Columns

$$S = \begin{pmatrix} \texttt{b} & \texttt{a} & \texttt{n} & \texttt{a} & \texttt{n} & \texttt{a} \\ \texttt{a} & * & \texttt{n} & * & \texttt{n} & * \\ \texttt{b} & * & * & \texttt{c} & \texttt{z} & \texttt{a} \\ \texttt{z} & \texttt{a} & \texttt{z} & * & \texttt{n} & * \\ \texttt{b} & \texttt{a} & * & \texttt{a} & \texttt{c} & \texttt{a} \\ \texttt{b} & * & \texttt{n} & \texttt{c} & * & \texttt{z} \end{pmatrix} \quad d = \begin{pmatrix} 4 \\ 3 \\ 4 \\ 2 \\ 4 \\ 3 \end{pmatrix} \quad d' = \begin{pmatrix} 4 \\ 2 \\ 4 \\ 1 \\ 4 \\ 3 \end{pmatrix}$$

$$T(n, \ell) = (\widetilde{d} + 1) T(n, \ell - 1) + n\ell$$

$$\implies T(n, \ell) = \mathcal{O}\left(\left(\widetilde{d} + 1\right)^{\ell} \cdot n\ell\right) = \mathcal{O}^*\left(\ell^{\ell}\right)$$

with $\widetilde{d} = \max_i d_i$

Linear Algorithm for $d_i \leq 1$

We solve ConRMC with $\max_i d_i \leq 1$ in $\mathcal{O}(n\ell)$

Linear Algorithm for $d_i \leq 1$

We solve ConRMC with $\max_i d_i \leq 1$ in $\mathcal{O}(n\ell)$

2-SAT

Possibly repeated literals $\{\alpha_i\}_{i=1\ldots n}$ and $\{\beta_i\}_{i=1\ldots n}$
Decide if the following is satisfiable

$$\bigwedge_{i=1}^{n} (\alpha_i \vee \beta_i)$$

2-SAT is solvable in $\mathcal{O}(n)$!

We solve ConRMC with $\max_i d_i \leq 1$ in $\mathcal{O}(n\ell)$

2-SAT

Possibly repeated literals $\{\alpha_i\}_{i=1\dots n}$ and $\{\beta_i\}_{i=1\dots n}$

Decide if the following is satisfiable

$$\bigwedge_{i=1}^{n} (\alpha_i \vee \beta_i)$$

2-SAT is solvable in $\mathcal{O}(n)$!

Reduction Time!

Linear Algorithm for $d_i \leq 1$

Suppose $v \in \Sigma^\ell$ solves MinRMC

$$x_{j,\sigma} = (\texttt{v[j] == } \sigma)$$

Linear Algorithm for $d_i \leq 1$

Suppose $v \in \Sigma^\ell$ solves MinRMC

$$x_{j,\sigma} = (\texttt{v[j] == } \sigma)$$

We write MinRMC as $\phi_1 \wedge \phi_2 \wedge \phi_3$ where:

$$\phi_1 = \text{“ For each } j \text{ at most one of } \{x_{j,\sigma}\}_{\sigma \in \Sigma} \text{ is true ”}$$

Linear Algorithm for $d_i \leq 1$

Suppose $v \in \Sigma^\ell$ solves MinRMC

$$x_{j,\sigma} = (\texttt{v[j] == } \sigma)$$

We write MinRMC as $\phi_1 \wedge \phi_2 \wedge \phi_3$ where:

$$\phi_1 = \text{" For each } j \text{ at most one of } \{x_{j,\sigma}\}_{\sigma \in \Sigma} \text{ is true "}$$

$$\phi_2 = \text{" For each } j \text{ if } d_i = 0 \text{ then } x_{j,S[i,j]} \text{ is true "}$$

Linear Algorithm for $d_i \leq 1$

Suppose $v \in \Sigma^\ell$ solves MinRMC

$$x_{j,\sigma} = (\texttt{v[j] == } \sigma)$$

We write MinRMC as $\phi_1 \wedge \phi_2 \wedge \phi_3$ where:

$$\phi_1 = \text{“ For each } j \text{ at most one of } \{x_{j,\sigma}\}_{\sigma \in \Sigma} \text{ is true ”}$$

$$\phi_2 = \text{“ For each } j \text{ if } d_i = 0 \text{ then } x_{j,S[i,j]} \text{ is true ”}$$

$$\phi_3 = \text{“ If } d_i = 1 \text{ then at most one of } \left\{x_{j,S[i,j]}\right\}_{j \leq \ell} \text{ is false ”}$$

Linear Algorithm for $d_i \leq 1$

$P = \{p_1 \ldots p_m\}$ propositions, exists a compact 2-SAT encoding of

$$C_{\leq 1}(P) = \text{“At most one proposition in } P \text{ is true”}$$

that uses $\mathcal{O}(m)$ clauses.

Linear Algorithm for $d_i \leq 1$

$P = \{p_1 \ldots p_m\}$ propositions, exists a compact 2-SAT encoding of

$$C_{\leq 1}(P) = \text{"At most one proposition in } P \text{ is true"}$$

that uses $\mathcal{O}(m)$ clauses.

Some more logic manipulations do the trick...

Open Problems

- MinRMC with t outliers: FPT w.r.t. $d + k + t$.
- Closest String is solved in $\mathcal{O}^*\left((16|\Sigma|)^d\right)$. Match that.

Thank You!