

Parameterized Algorithms for Matrix Completion with Radius Constraints

Tomohiro Koana¹ Vincent Froese¹
Rolf Niedermeier¹

¹Technische 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 $*$

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

Problem Statement

For $i = 1 \dots n$:

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

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

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. $\forall i \delta(v, T[i]) \leq d$?

Problem Statement

For $i = 1 \dots n$:

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

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

Constraint Radius Matrix Completion (ConRMC)

Input: Matrix $S \in (\Sigma \cup \{*\})^{n \times \ell}$ and $d_1 \dots d_n \in \mathbb{N}$.

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

Stringology and ML meet in the middle!

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)

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

All Problems are Intractable

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

All Problems are Intractable

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

Fixed Parameter Tractability

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

All Problems are Intractable

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

Fixed Parameter Tractability

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

CPM '14 solved MinRMC in:

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

Here we solve MinRMC in:

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

CPM '14 solved MinRMC in:

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

Here we solve MinRMC in:

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

k is the maximum number of character row-wise

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

Parameter: Number ℓ of Columns

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

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

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

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

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

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

If holds:

return True

Else:

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

Parameter: Number ℓ of Columns

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

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:

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

Parameter: Number ℓ of Columns

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

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

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

Parameter: Number ℓ of Columns

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

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

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

For $j \in R_i$:

Set $v[j] = S[i, j]$, process j -th column and recurse

Parameter: Number ℓ of Columns

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

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

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

For $j \in R_i$:

Set $v[j] = S[i, j]$, process j -th column and recurse

Parameter: Number ℓ of Columns

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

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

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

For $j \in R_i$:

Set $v[j] = S[i, j]$, process j -th column and recurse

Parameter: Number ℓ of Columns

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

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

$$\text{with } \tilde{d} = \max_i d_i$$

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 $\mathcal{O}(nl)$

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 $\mathcal{O}(n)$!

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 $\{\alpha_i\}_{i=1\dots n}$ 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 ConRMC

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

Linear Algorithm for $d_i \leq 1$

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

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

We write ConRMC 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 ConRMC

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

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

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

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

Linear Algorithm for $d_i \leq 1$

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

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

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

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

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

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

Linear Algorithm for $d_i \leq 1$

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

“At most one proposition in $\{p_1 \dots p_m\}$ is true”

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

Linear Algorithm for $d_i \leq 1$

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

“At most one proposition in $\{p_1 \dots p_m\}$ is true”

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

Some more logic manipulations do the trick...

- 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!