SUPPLEMENTARY MATERIAL

Wenzhe Lu and Heng Qiao

Appendix A
Proof of Theorem 1

In the worst scenario, the upper bound $\mathcal{B}_u$ from the warm start is not exploited to prune the tree and all possible cases will be visited during the search. Thus, all the indices in the range $[0, N-1]$ can be chosen as the branching index in the B&B algorithm (see Step 11 in Table 1). In our proof, for ease of exposition, we assume that the branching indices are chosen in an ascending order. To calculate the total number of nodes, we enumerate all possible cases with respect to the size of $\mathbb{S}_1$ at each node, i.e., $|\mathbb{S}_1|$.

A. *Number of nodes with* $|\mathbb{S}_1| = s$

When $|\mathbb{S}_1| = s$, no more indices are left undetermined, i.e., $\overline{\mathbb{S}} = \varnothing$ and $\mathbb{S}_0 = \mathbb{S}_1^{\mathrm{C}} = \{0, \cdots, N-1\}\backslash\mathbb{S}_1$. So we only care about all possibilities of $\mathbb{S}_1$ satisfying Δ-separation.

We first observe that there are $(s+1)$ blocks of zeros. Those blocks satisfy four constraints: blocks at head and tail must be at least of zero size; interior blocks must be at least of size Δ; the sum of the sizes of blocks at head and tail must be at least Δ; the sum of the sizes of all $(s+1)$ blocks is $(N-s)$. Assuming X_i $(1 \leqslant i \leqslant s+1)$ denotes the length of the corresponding block of zeros, then the following relationship holds:

$$\begin{aligned}
\textstyle\sum_{i=1}^{s+1} X_i &= N - s\\
X_1, X_{s+1} &\geqslant 0\\
X_1 + X_{s+1} &\geqslant \Delta\\
X_i &\geqslant \Delta,\ \forall i \in \{2, \dots, s\}.
\end{aligned}$$

Now, we simplify this relationship according to X_1. When $X_1 \in \{0, \dots, \Delta-1\}$, the relationship becomes

$$\begin{aligned}
\textstyle\sum_{i=2}^{s+1} X_i &= N - s - X_1\\
X_{s+1} &\geqslant \Delta - X_1\\
X_i &\geqslant \Delta,\ \forall i \in \{2, \dots, s\}.
\end{aligned}$$

Based on combinatorics, $\forall X_1 \in \{0, \dots, \Delta-1\}$, the number of ways to assign $(s+1)$ blocks is $C_{N-s\Delta-1}^{s-1}$. So considering all Δ cases of X_1, the total number of ways to assign $(s+1)$ blocks is $\Delta \cdot C_{N-s\Delta-1}^{s-1}$.

Otherwise, when $X_1 \in \{\Delta, \dots, N-s-(s-1)\Delta\}$, the relationship becomes

$$\begin{aligned}
\textstyle\sum_{i=2}^{s+1} X_i &= N - s - X_1\\
X_{s+1} &\geqslant 0\\
X_i &\geqslant \Delta,\ \forall i \in \{2, \dots, s\}.
\end{aligned}$$

Based on combinatorics, $\forall X_1 \in \{\Delta, \dots, N-s-(s-1)\Delta\}$, the number of ways to assign $(s+1)$ blocks is $C_{N-(s-1)\Delta-1-X_1}^{s-1}$. So considering all cases here, the total number of ways to assign $(s+1)$ blocks is $\sum_{i=s}^{N-s\Delta} C_{i-1}^{s-1}$.

Thus, considering all possibilities of X_1, the total number of ways to assign blocks under $|\mathbb{S}_1| = s$ is

$$\begin{aligned}
\beta^s &= \Delta \cdot C_{N-s\Delta-1}^{s-1} + \textstyle\sum_{i=s}^{N-s\Delta} C_{i-1}^{s-1}\\
&= \Delta \cdot C_{N-s\Delta-1}^{s-1} + C_{N-s\Delta}^{s} \qquad \text{(Hockey-stick identity)}\\
&= \Delta \cdot C_{N-s\Delta-1}^{s-1} + \frac{N-s\Delta}{s} C_{N-s\Delta-1}^{s-1}\\
&= \frac{N}{s} C_{N-s\Delta-1}^{s-1}.
\end{aligned}$$

B. *Number of nodes with* $|\mathbb{S}_1| = s-1$

When $|\mathbb{S}_1| = s-1$, the situation is much more complicated since particular attention needs to be paid to the assignment of $\mathbb{S}_0$ and $\overline{\mathbb{S}}$. Notice that we choose the branching indices in ascending order, so the indices smaller than the largest index in $\mathbb{S}_1$ must belong to $\mathbb{S}_1$ or $\mathbb{S}_0$. Moreover, a total of *at least* Δ indices belong to $\mathbb{S}_0$ before the smallest index and after the largest index in $\mathbb{S}_1$ because of wrap-around separation. And the left indices need to be considered for assignment in $\mathbb{S}_0$ and $\overline{\mathbb{S}}$.

So it is a natural choice to sort by the number of indices which have already been decided in $\mathbb{S}_1$ or $\mathbb{S}_0$. Let us first consider the smallest possible value γ of this number. When $|\mathbb{S}_1| = s-1$, there are s blocks of zeros. Assuming X_i $(1 \leqslant i \leqslant s)$ denotes the length of the corresponding block of zeros, we clearly find that the smallest possible value can be achieved when $X_1 + X_s = \Delta$ and $X_i = \Delta$ $(\forall i \in \{2, \dots, s-1\})$. Thus,

$$\gamma = (s-1) + (s-1)\Delta = (s-1)(\Delta+1).$$

Now we focus on the case where the smallest $\gamma = (s-1)(\Delta+1)$ indices have been decided in $\mathbb{S}_1$ or $\mathbb{S}_0$. According to the result of last section, the number of ways to assign blocks under $|\mathbb{S}_1| = s-1$ among $\gamma = (s-1)(\Delta+1)$ indices is $\frac{\gamma}{s-1} C_{\gamma-(s-1)\Delta-1}^{s-2} = \frac{\gamma}{s-1} C_{s-2}^{s-2}$.

However, it doesn't simply mean the left $(N-\gamma)$ indices are undecided. For example, let $N = 10$, $s = 3$, $\Delta = 2$. Then $\frac{\gamma}{s-1} C_{s-2}^{s-2} = 3$ ways to assign blocks are listed below

$$100100xx\boxed{00},\ 010010\boxed{0}xx\boxed{0},\ 001001\boxed{00}xx,$$

where 1, 0 and x denote the index decided in $\mathbb{S}_1$, $\mathbb{S}_0$ and undecided respectively. Notice that 0s in the boxes also belong to $\mathbb{S}_0$ because of wrap-around separation. That is to say, if $\gamma \leqslant N-\Delta$, $(N-\gamma-\Delta)$ indices remain to be assigned in $\mathbb{S}_0$ or $\overline{\mathbb{S}}$. If $N-\Delta+1 \leqslant \gamma$, no index remains undecided. For $(N-\gamma-\Delta)$ undecided indices, there are $(N-\gamma-\Delta+1)$ ways to assign them in $\mathbb{S}_0$ or $\overline{\mathbb{S}}$ because the branching indices are chosen in ascending order, which means all indices assigned in $\mathbb{S}_0$ must be smaller than all indices assigned in $\overline{\mathbb{S}}$, e.g.,

$$100100xx\boxed{00} \rightarrow \begin{cases} 100100xx\boxed{00} \\ 1001000x\boxed{00} \\ 10010000\boxed{00}. \end{cases}$$

Similarly, for zero undecided index, there is only 1 way to assign them in $\mathbb{S}_0$ or $\overline{\mathbb{S}}$. That is to say, for $|\mathbb{S}_1| = s-1$ and $\gamma = (s-1)(\Delta+1)$, the total number of ways to assign all N indices is

$$\beta^{s-1,\gamma} = \begin{cases} \frac{\gamma}{s-1} C_{s-2}^{s-2} \cdot (N-\gamma-\Delta+1) & \text{, if } \gamma \leqslant N-\Delta \\ \frac{\gamma}{s-1} C_{s-2}^{s-2} & \text{, otherwise.} \end{cases}$$

Next, we consider the case that the smallest $(\gamma+1)$ indices have been decided in $\mathbb{S}_1$ or $\mathbb{S}_0$. The number of ways to assign blocks under $|\mathbb{S}_1| = s-1$ among $(\gamma+1)$ indices is $\frac{\gamma+1}{s-1} C_{s-1}^{s-2}$.

At this time, the undecided indices are dependent on the block where the 0 is added when $\gamma \rightarrow \gamma+1$. Still considering the setting where $N = 10$, $s = 3$, $\Delta = 2$. Then $\frac{\gamma+1}{s-1} C_{s-1}^{s-2} = 7$ ways to assign blocks are listed below

$$\begin{cases} (1)\ 100100\underline{0}x\boxed{00},\ 010010\underline{0}xx\boxed{0},\ 001001\underline{0}\boxed{0}xx, \\ (2)\ 100\underline{0}100x\boxed{00},\ 0100\underline{0}10\boxed{0}x\boxed{0},\ 001000\underline{0}1\boxed{00}x, \\ (3)\ \underline{0}001001\boxed{00}x, \end{cases}$$

where underlined 0s are the added 0s when $\gamma \rightarrow \gamma+1$. For case (1), the 0 is added to the tail zero block. The ways of assignment in case (1) have been counted in $\beta^{s-1,\gamma}$ since the smallest γ indices remain the same. For cases (2) and (3), the 0 is either added to the interior block or added to the head block when the γ-th index is 1. Under these cases, wrap-around separation shown by boxes will not be effected. Therefore, the ways of assignment in cases (2) and (3) which have not been counted before can be calculated as

$$\beta^{s-1,\gamma+1} = \begin{cases} \left(\frac{\gamma+1}{s-1} C_{s-1}^{s-2} - \frac{\gamma}{s-1} C_{s-2}^{s-2}\right)(N-\gamma-\Delta) \\ \qquad\qquad\qquad\qquad\qquad\qquad \text{, if } \gamma+1 \leqslant N-\Delta \\ \left(\frac{\gamma+1}{s-1} C_{s-1}^{s-2} - \frac{\gamma}{s-1} C_{s-2}^{s-2}\right), \text{ otherwise.} \end{cases}$$

Considering all cases from γ to N, the total number of ways to assign N indices under $|\mathbb{S}_1| = s-1$ is

$$\begin{aligned} \beta^{s-1} = & \frac{\gamma}{s-1} C_{s-2}^{s-2} \cdot (N-\gamma-\Delta+1) \\ & + \left(\frac{\gamma+1}{s-1} C_{s-1}^{s-2} - \frac{\gamma}{s-1} C_{s-2}^{s-2}\right)(N-\gamma-\Delta) + \dots \\ & + \left(\frac{N-\Delta}{s-1} C_{N-s\Delta-1}^{s-2} - \frac{N-\Delta-1}{s-1} C_{N-s\Delta-2}^{s-2}\right) + \dots \\ & + \left(\frac{N}{s-1} C_{N-(s-1)\Delta-1}^{s-2} - \frac{N-1}{s-1} C_{N-(s-1)\Delta-2}^{s-2}\right). \end{aligned}$$

After we rearrange and remove the duplicates,

$$\beta^{s-1} = \sum_{t=(s-1)(\Delta+1)}^{N-\Delta-1} \frac{t}{s-1} C_{t-(s-1)\Delta-1}^{s-2} + \frac{N}{s-1} C_{N-(s-1)\Delta-1}^{s-2}.$$

C. Total number of nodes

For $|\mathbb{S}_1| \in \{1, \dots, s-2\}$, the analysis is similar to that in the previous section. For $|\mathbb{S}_1| = 0$, all N indices are undetermined, that is, can be designated either to 1 or 0. As we assume that the chosen branching indices are in an ascending order, there are $(N+1)$ ways to assign them to $\mathbb{S}_0$ or $\overline{\mathbb{S}}$.

In summary, the total number of nodes under separation Δ can be calculated as

$$\beta_\Delta = \sum_{i=1}^{s} \frac{N}{i} C_{N-i\Delta-1}^{i-1} + \sum_{i=1}^{s-1} \left(\sum_{t=i(\Delta+1)}^{N-\Delta-1} \frac{t}{i} C_{t-i\Delta-1}^{i-1} \right) + C_{N+1}^{1}.$$

D. Lower bound of the node reduction

We first consider the case without separation, i.e., $\Delta = 0$,

$$\begin{aligned} \beta_0 &= \sum_{i=1}^{s} \frac{N}{i} C_{N-1}^{i-1} + \sum_{i=1}^{s-1} \left(\sum_{t=i}^{N-1} \frac{t}{i} C_{t-1}^{i-1} \right) + C_{N+1}^{1} \\ &= \sum_{i=1}^{s} C_N^i + \sum_{i=1}^{s-1} \left(\sum_{t=i}^{N-1} C_t^i \right) + C_{N+1}^{1} \\ &= C_N^s + \sum_{i=1}^{s-1} \left(\sum_{t=i}^{N} C_t^i \right) + C_{N+1}^{1} \\ &= C_N^s + \sum_{i=0}^{s-1} C_{N+1}^{i+1}. \end{aligned}$$

So our goal is to provide a lower bound of the node reduction ($\beta_{\text{diff}} = \beta_0 - \beta_\Delta$) caused by separation Δ. To achieve this goal, we first derive an upper bound of β_Δ by considering straight separation, i.e.,

$$\forall a \neq b \in \mathbb{S},\ \varphi(a,b) = |a-b| > \Delta.$$

From the definition, it is clear that $\varphi(m,n) \geqslant \phi(m,n)$, so the total number of nodes under straight separation $\overline{\beta} \geqslant \beta_\Delta$.

To calculate $\overline{\beta}$, we still classify all nodes by $|\mathbb{S}_1|$. When $|\mathbb{S}_1| = s$, we only care about all possibilities of $\mathbb{S}_1$ satisfying straight separation. Assuming X_i $(1 \leqslant i \leqslant s+1)$ denotes the length of the corresponding block of zeros, then the following relationship holds:

$$\begin{aligned} &\textstyle\sum_{i=1}^{s+1} X_i = N-s \\ &X_1, X_{s+1} \geqslant 0 \\ &\qquad X_i \geqslant \Delta,\ \forall i \in \{2, \dots, s\}. \end{aligned}$$

Based on combinatorics, the total number of ways to assign $(s+1)$ blocks is $C_{N-(s-1)\Delta}^{s}$.

For $|\mathbb{S}_1| \in \{0, \dots, s-1\}$, we follow the idea in the context of wrap-around separation. Then the total number of nodes under straight separation is calculated as

$$\overline{\beta} = C_{N-(s-1)\Delta}^{s} + \sum_{i=0}^{s-1} \left(C_{N-i\Delta}^{i+1} + C_{N-(i-1)\Delta}^{i} \right).$$

Thus, we can bound the node reduction as

$$
\begin{aligned}
&\beta_{\text{diff}} = \beta_0 - \beta_\Delta \geqslant \beta_0 - \overline{\beta} \\
&= C_N^s - C_{N-(s-1)\Delta}^s + \sum_{i=1}^{s-1}(C_{N+1}^{i+1} - C_{N-i\Delta}^{i+1} - C_{N-(i-1)\Delta}^{i}) \\
&\geqslant C_N^s - C_{N-(s-1)\Delta}^s + \sum_{i=1}^{s-1}(C_{N+1}^{i+1} - C_{N-(i-1)\Delta+1}^{i+1}).
\end{aligned}
$$

Since $\frac{N-(s-1)\Delta-k}{N-k}$ is monotonically decreasing with $k \in [0, s-1]$, we have

$$
\begin{aligned}
\frac{C_{N-(s-1)\Delta}^s}{C_N^s} &= \frac{[N-(s-1)\Delta]\cdots[N-(s-1)(\Delta+1)]}{N\cdots(N-s+1)} \\
&\leqslant (\frac{N-(s-1)\Delta}{N})^s.
\end{aligned}
$$

Thus, we can further bound the reduction as

$$
\begin{aligned}
&\beta_{\text{diff}} \\
&\geqslant \left[1-\left(1-\frac{(s-1)\Delta}{N}\right)^s\right]C_N^s + \sum_{i=1}^{s-1}\left[1-\left(1-\frac{(i-1)\Delta}{N+1}\right)^{i+1}\right]C_{N+1}^{i+1} \\
&\geqslant \left[1-\left(1-\frac{(s-1)\Delta}{N}\right)^s\right]C_N^s + \left[1-\left(1-\frac{(s-2)\Delta}{N+1}\right)^s\right]C_{N+1}^s.
\end{aligned}
$$

B Complete Pseudo Code of Sep-CoSaMP

Table 1 Sep-CoSaMP

Input: $\{\mathbf{y}[l]\}_{l=1}^L$, $\mathbf{A}$, Δ, $\mathcal{M}$, s and max_itr. Mapping $\mathbb{M}$. Define $\mathbf{r}$ and $\widetilde{\mathbf{A}}$ as for (Co-MIP).

```
i = 0, p_i = 0, e = r.
while i < max_itr do
    i ← i + 1
    s ← Ã^T e
    Ω ← Supp (s_𝕄)
    Ω ← Ω ⋃ Supp (p_{i−1})
    b_Ω ← Ã_Ω^† r, b_{Ω^c} ← 0
    p_i ← b_𝕄
    e ← r − Ã p_i
end while
```

Output: The support estimate $\mathbb{S}^{\#} = \text{Supp}(\mathbf{p}_i)$ and the corresponding objective function value $\mathcal{B}_u = \frac{1}{2}||\mathbf{r} - \widetilde{\mathbf{A}}\mathbf{p}_i||_2^2$

Notation: † denotes Moore-Penrose inverse. Subscript $\mathbb{M}$ represents the image under mapping $\mathbb{M}$. $\mathbf{b}_\Omega$ and $\widetilde{\mathbf{A}}_\Omega$ represent the sub-vector of $\mathbf{b}$ with support Ω and the sub-matrix of $\widetilde{\mathbf{A}}$ obtained by keeping the column vectors in Ω respectively.