A APPENDIX

A.1 D_∞ — A TIGHTER RESULT FOR LOG-SUPERMODULAR MODELS

In the main paper, we prove that clamping always improve D_∞ for general models, and this naturally implies the same result for log-supermodular models. In that proof, we derive the fact that $A_l \geq \widehat{\mathcal{Z}}_l$. Here we provide an auxiliary proof for log-supermodular models, and the purpose of this proof is to stress a stronger result in this special case, *i.e.*, $A_l = \widehat{\mathcal{Z}}_l$. We consider multi-label log-supermodular models for D_∞. In this case, one relaxation of D_∞ reads as following [23],

$$\log \widehat{\mathcal{Z}} = \min_{\mathbf{s}\in B(f)} \sum_{i=1}^{N} \log \sum_{j=1}^{L} \exp(-s_{i,j}),$$

where $B(f) = \{\mathbf{s} \mid \forall x \in \{0,1\}^{NL} : \mathbf{x}^T\mathbf{s} \leq f(x); \mathbf{1}^T\mathbf{s} = f(\mathbf{1})\}$ is the well-known base polytope.

$$\begin{aligned} \mathcal{Z}(f) &= \sum_{\mathbf{x}\in\mathcal{X}} \exp(-f(\mathbf{x})) \\ &= \sum_{l=1}^{L} \sum_{\mathbf{x}\in\mathcal{X}, \mathbf{x}_{k,l}=1} \exp(-f(\mathbf{x})) \\ &= \sum_{l=1}^{L} \underbrace{\sum_{\mathbf{x}\in\mathcal{X}_{-k}} \exp(-f(\mathbf{x}+\mathbf{e}_{k,l}))}_{\mathcal{Z}(f_{+k,l})} \end{aligned} \tag{7}$$

We can obtain the upper bound of $\mathcal{Z}(f_{+k,l})$, denoted by $\widehat{\mathcal{Z}}_l$, using L-FIELD. Because submodular functions are closed under clamping, i.e. f is submodular so are $f_{+k,l}$ (see e.g. [24])[2]. Obviously $\mathcal{Z}(f) = \sum_l \mathcal{Z}(f_{+k,l}) \leq \sum_l \widehat{\mathcal{Z}}_l$, hence $\sum_l \widehat{\mathcal{Z}}_l$ is an upper bound for the partition function.

Theorem A.1. *For multi-label log-supermodular models, i.e., when f is submodular, we have $\sum_l \widehat{\mathcal{Z}}_l \leq \widehat{\mathcal{Z}}$, where k is of arbitrary choice.*

Following the same path, we first decompose the objective used in the optimization problem for solving $\widehat{\mathcal{Z}}$ as follows:

$$\begin{aligned} \exp\Big(\sum_{i=1}^{N} \log \sum_{l=1}^{L} e^{-s_{i,l}}\Big) &= \prod_{i=1}^{N}\Big(\sum_{l=1}^{L} e^{-s_{i,l}}\Big) \\ &= \sum_{l=1}^{L}\Big(e^{-s_{k,l}} \cdot \prod_{i\neq k}\Big(\sum_{j=1}^{L} e^{-s_{i,j}}\Big)\Big) \end{aligned} \tag{8}$$

[2] Typically $f_{+k,l}$ is defined as $f_{+k,l}(\mathbf{x}) = f(\mathbf{x}+\mathbf{e}_{k,l}) - f(\mathbf{e}_{k,l})$ to make sure that $f_{+k,l}$ is normalized as $f_{+k,l}(\mathbf{0}) = 0$, but this is of course w.l.o.g.

Next we have $A_l = \min_{s\in B(f)} e^{-s_{k,l}} \cdot \prod_{i\neq k}(\sum_{j=1}^{L} e^{-s_{i,j}})$, then it trivially follows that

$$\begin{aligned} \widehat{\mathcal{Z}} &= \min_{s\in B(f)} \sum_{l=1}^{L}\Big(e^{-s_{k,l}} \cdot \prod_{i\neq k}\Big(\sum_{j=1}^{L} e^{-s_{i,j}}\Big)\Big) \\ &\geq \sum_{l=1}^{L} \min_{s\in B(f)}\Big(e^{-s_{k,l}} \cdot \prod_{i\neq k}\Big(\sum_{j=1}^{L} e^{-s_{i,j}}\Big)\Big) \\ &= \sum_{l} A_l \end{aligned} \tag{9}$$

The following lemma says that $\forall l, A_l = \widehat{\mathcal{Z}}_l$, hence $\widehat{\mathcal{Z}} \geq \sum_l A_l = \sum_l \widehat{\mathcal{Z}}_l$. Note this is a stronger result compared to the proof for the general case.

Lemma A.2. $A_l = \widehat{\mathcal{Z}}_l$.

Proof. This is equivalent to proving that $\log A_l = \log \widehat{\mathcal{Z}}_l$. Later we will show that $\log A_l$, the minimum of $-s_{k,l} + \sum_{i\neq k} \log \sum_{j=1}^{L}(1 + e^{-s_{i,j}})$ for $s \in B(f)$, can still be achieved if we fix $s_{k,l} = f(\mathbf{e}_{k,l})$. If this is true, we can replace $s_{k,l}$ with $f(\mathbf{e}_{k,l})$ in $B(f)$ and get the following explicit form.

$$\begin{cases} \forall \mathbf{x} \in \{0,1\}^{NL}, \mathbf{x}_{k,l} = 0, \mathbf{x}^T s + f(\mathbf{e}_{k,l}) \leq f(\mathbf{x}+\mathbf{e}_{k,l}) \\ \forall \mathbf{x} \in \{0,1\}^{NL}, \mathbf{x}_{k,l} = 0, \mathbf{x}^T s \leq f(\mathbf{x}) \\ (\mathbf{1}-\mathbf{e}_{k,l})^T s + f(\mathbf{e}_{k,l}) = f(\mathbf{1}) \end{cases}$$

Notice that the second constraint is redundant, because the first inequality requires $\forall \mathbf{x} \in \{0,1\}^N, \mathbf{x}_{k,l} = 0, \mathbf{x}^T s \leq f(\mathbf{x}+\mathbf{e}_{k,l}) - f(\mathbf{e}_{k,l})$ and $f(\mathbf{x}+\mathbf{e}_{k,l}) - f(\mathbf{e}_{k,l}) \leq f(\mathbf{x})$ by submodularity, and for the same reason the last equality fulfills the second inequality when $\mathbf{x} = \mathbf{1} - \mathbf{e}_{k,l}$. Thus we can remove the second constraint in above inequality system.

Now we write the explicit form of $B(f_{+k,l})$ as follows.

$$\begin{cases} \forall \mathbf{x} \in \{0,1\}^{NL}, \mathbf{x}_{v_{k,l}} = 0, \mathbf{x}^T s \leq f(\mathbf{x}+\mathbf{e}_{k,l}) - f(\mathbf{e}^{v_{k,l}}) \\ (\mathbf{1}-\mathbf{e}_{k,l})^T s = f(\mathbf{1}) - f(\mathbf{e}_{k,l}) \end{cases}$$

Observe that this is the same as $B(f) \cap \{s \mid s_{k,l} = f(\mathbf{e}_{k,l})\}$. Hence the feasible regions of two minimization problem are exactly the same. Furthermore, since we fix $s_{k,l} = f(\mathbf{e}_{k,l})$, the objective of $\log A_l$ changes into $-f(\mathbf{e}_{k,l}) + \sum_{i\neq k} \log \sum_{j=1}^{L}(1+e^{-s_{i,j}})$, which is also the same as the objective of $\widehat{\mathcal{Z}}_l$. Therefore, $\log A_l = \log \widehat{\mathcal{Z}}_l$, which implies $A_l = \widehat{\mathcal{Z}}_l$. □

Lemma A.3. *By adding $s_{k,l} = f(\mathbf{e}_{k,l})$ to the constraint set, the optimal value of the optimization problem $\min_{s\in B(f)} -s_{k,l} + \sum_{i\neq k} \log \sum_{j=1}^{L}(1+e^{-s_{i,j}})$ does not change.*

Proof. First we define $g(s) = \sum_{i \neq k} \log \sum_{j=1}^{L}(1 + e^{-s_{i,j}}) - s_{k,l}$. Then we have

$$\begin{cases} \frac{\partial g}{\partial s_{k,l}} = -1 \\ \frac{\partial g}{\partial s_{i,j}} = \frac{-\exp(-s_{i,j})}{\sum_{j'=1}^{L}(1+\exp(-s_{i,j'}))} > -1, \forall v_{i,j} \in V \setminus \{v_{k,l}\} \end{cases}$$

Hence it is easy to see exchange $\Delta \geq 0$ between $s_{i,j}$ and $s_{k,l}$, i.e. $s'_{k,l} = s_{k,l} + \Delta, s'_{i,j} = s_{i,j} - \Delta$, can only decrease the objective. Therefore, given optimal solution s^*, we can get a solution at least as good as s^* by setting $s'_{k,l} = s^*_{k,l} + \Delta$ and $s'_{i,j} = s^*_{i,j} - \Delta$ for arbitrary $(i,j) \neq (k,l)$. We can exploit this property to change $s^*_{k,l}$ into $f(\mathbf{e}_{k,l})$, but we need to guarantee that every exchange results in a feasible solution. Hence we need to deal with exchange capacity $\hat{c}(s;(k,l),e') = \min\{f(\mathbf{1}_A) - \mathbf{1}_A^T\mathbf{s}, \forall A \supseteq (k,l), e' \notin A\}$ (e' denotes the element-label pair to exchange with (k,l), $\mathbf{1}_A$ is the indicator vector for set A). Let $S_{e'} \cup \{(k,l)\}$ be the set that achieves $\hat{c}(s;(k,l),e')$, we know $e' \notin S_{e'} \cup \{(k,l)\}$. We propose the following procedure *exchange*, and we will prove later this algorithm will make $s'_{k,l} = f(\mathbf{e}_{k,l})$. Since we already proved that exchange always results in better solution, this will finish the proof.

procedure *exchange*():
Initiate $U = [N] \times [L] \setminus \{(k,l)\}, s = s^*$;
While $U \neq \emptyset$:
  Arbitrarily pick $e' \in U$;
  $s_{k,l} \leftarrow s_{k,l} + \hat{c}(s;(k,l),e')$;
  $s_{e'} \leftarrow s_{e'} - \hat{c}(s;(k,l),e')$;
  $U = U \cap S_{e'}$
end;

We first show that after one exchange with e' the new modular function s' is tight at $S_{e'} \cup \{(k,l)\}$, *i.e.*, $\mathbf{1}^T_{S_{e'}\cup\{(k,l)\}} s' = f(\mathbf{1}_{S_{e'}\cup\{(k,l)\}})$.

$$\begin{aligned} s'_{k,l} &= s_{k,l} + \hat{c}(s;(k,l),e') \\ &= s_{k,l} + f(\mathbf{1}_{S_{e'}\cup\{(k,l)\}}) - \mathbf{1}^T_{S_{e'}\cup\{(k,l)\}} s \\ &= s_{k,l} + f(\mathbf{1}_{S_{e'}\cup\{(k,l)\}}) - \mathbf{1}^T_{S_{e'}} s - s_{k,l} \\ &= f(\mathbf{1}_{S_{e'}\cup\{(k,l)\}}) - \mathbf{1}^T_{S_{e'}} s \\ \Rightarrow & \mathbf{1}^T_{S_{e'}\cup\{(k,l)\}} s' = f(\mathbf{1}_{S_{e'}\cup\{(k,l)\}}) \end{aligned} \quad (10)$$

Because s' is tight at $S_{e'} \cup \{(k,l)\}$, the element picked next round must be the element in $S_{e'}$ such that the next exchange also results in a feasible solution, otherwise the next exchange will break the tight upper bound for s' at $S_{e'} \cup \{(k,l)\}$ since we only increase $s'_{k,l}$. This is why we let $U = U \cap S_{e'}$ in the algorithm. It is also obvious that once s' is tight at $S_{e'} \cup \{(k,l)\}$, it will always be tight at $S_{e'} \cup \{(k,l)\}$. Moreover, notice that $e' \notin S_{e'}$ but $e' \in U$, hence the intersection operation always strictly decreases the size of U in each round. Therefore, algorithm will terminate and U will definitely turn into $\emptyset$. The final U is $\cap_{e'} S_{e'}$, hence $\cap_{e'}(S_{e'} \cup \{(k,l)\}) = (\cap_{e'} S_{e'}) \cup \{(k,l)\} = \{(k,l)\}$. Since the final s' is tight at each $S_{e'} \cup \{(k,l)\}$ and it is well-known result that the intersection of tight sets is also tight. Therefore the final s' is tight at $\{(k,l)\}$, i.e. $s'_{k,l} = f(\mathbf{e}_{k,l})$, which completes the proof. □

A.2 THE LOWER BOUND FOR BINARY LOG-SUPERMODULAR MODELS: CLAMPING ALWAYS HELPS

Moreover, in [6], the authors also use the properties of submodular functions to obtain lower bounds on $\mathcal{Z}$ for binary log-supermodular models. Likewise, clamping can only improve the lower bounds.

Since we are only concerned with binary models, we change to the set notation for energy function to be in line with [6]. We denote the energy function by $F : 2^V \to \mathbb{R}^n$, where $V = \{1, \cdots, N\}$ is the ground set of all the elements. We notate two operations that preserve submodularity, (a) **contraction**: $F_X(A) = F(X \cup A) - F(X), A \subseteq V \setminus X$, (b) **restriction**: $F^X(A) = F(A), A \subseteq X$.

From the proof of Lemma 4 in [6], we know that by optimizing over bar supergradient, we can derive a lower bound of the log-partition function as follows,

$$\begin{aligned} \log \widehat{\mathcal{Z}}_L(F) = \max_{X \in V} -F(X) &+ \sum_{j \in X} \log(1 + e^{F(V) - F(V \setminus \{j\})}) \\ &+ \sum_{j \notin X} \log(1 + e^{-F(\{j\})}). \end{aligned} \quad (11)$$

After clamping i, we also apply this method to get the lower bound for $\mathcal{Z}(F_{\{i\}})$ and $\mathcal{Z}(F^{V\setminus\{i\}})$, denote them by $\widehat{\mathcal{Z}}_L(F_{\{i\}})$ and $\widehat{\mathcal{Z}}_L(F^{V\setminus\{i\}})$ respectively. It is obvious that $\mathcal{Z}(F) = e^{-F(\{i\})} \cdot \mathcal{Z}(F_{\{i\}}) + \mathcal{Z}(F^{V\setminus\{i\}}) \geq e^{-F(\{i\})} \cdot \widehat{\mathcal{Z}}_L(F_{\{i\}}) + \widehat{\mathcal{Z}}_L(F^{V\setminus\{i\}})$, so if $\widehat{\mathcal{Z}}_L(F) \leq e^{-F(\{i\})} \cdot \widehat{\mathcal{Z}}_L(F_{\{i\}}) + \widehat{\mathcal{Z}}_L(F^{V\setminus\{i\}})$, then $e^{-F(\{i\})} \cdot \widehat{\mathcal{Z}}_L(F_{\{i\}}) + \widehat{\mathcal{Z}}_L(F^{V\setminus\{i\}})$ is a better lower bound.

Theorem A.4. $\widehat{\mathcal{Z}}_L(F) \leq e^{-F(\{i\})} \cdot \widehat{\mathcal{Z}}_L(F_{\{i\}}) + \widehat{\mathcal{Z}}_L(F^{V\setminus\{i\}})$.

Proof. We take the exponent of $\log \widehat{\mathcal{Z}}_L(F)$, then $\widehat{\mathcal{Z}}_L(F) = \max_{X \in V} e^{-F(X)} \prod_{j \in X}(1 + e^{F(V)-F(V\setminus\{j\})}) \prod_{j \notin X}(1 + e^{-F(\{j\})})$. We split it into two cases. First, if $i \in X^*$, where X^* is the optimal element set for bar supergradient, we know

$$\begin{aligned}
\widehat{\mathcal{Z}}_L(F) = \max_{X\in V} e^{-F(X)} & \prod_{j\in X\setminus\{i\}} (1+e^{F(V)-F(V\setminus\{j\})}) \cdot \\
& (1+e^{F(V)-F(V\setminus\{i\})}) \prod_{j\notin X} (1+e^{-F(\{j\})}) \\
& = \max_{X\in V} e^{-F(X)} \prod_{j\in X\setminus\{i\}} (1+e^{F(V)-F(V\setminus\{j\})}) \cdot \\
& \prod_{j\notin X} (1+e^{-F(\{j\})}) + e^{F(V)-F(V\setminus\{i\})-F(X)} \cdot \\
& \prod_{j\in X\setminus\{i\}} (1+e^{F(V)-F(V\setminus\{j\})}) \prod_{j\notin X} (1+e^{-F(\{j\})}) \\
& := \max_{X\in V} A_1 + B_1.
\end{aligned} \tag{12}$$

Otherwise, if $i \notin X^*$

$$\begin{aligned}
\widehat{\mathcal{Z}}_L(F) = \max_{X\in V\setminus\{i\}} & e^{-F(X)} \prod_{j\in X} (1+e^{F(V)-F(V\setminus\{j\})}) \cdot \\
& \prod_{j\notin X\cup\{i\}} (1+e^{-F(\{j\})})(1+e^{-F(\{i\})}) \\
& = \max_{X\in V\setminus\{i\}} e^{-F(X)} \prod_{j\in X} (1+e^{F(V)-F(V\setminus\{j\})}) \cdot \\
& \prod_{j\notin X\cup\{i\}} (1+e^{-F(\{j\})}) + e^{-F(X)-F(\{i\})} \cdot \\
& \prod_{j\in X} (1+e^{F(V)-F(V\setminus\{j\})}) \prod_{j\notin X\cup\{i\}} (1+e^{-F(\{j\})}) \\
& := A_2 + B_2.
\end{aligned} \tag{13}$$

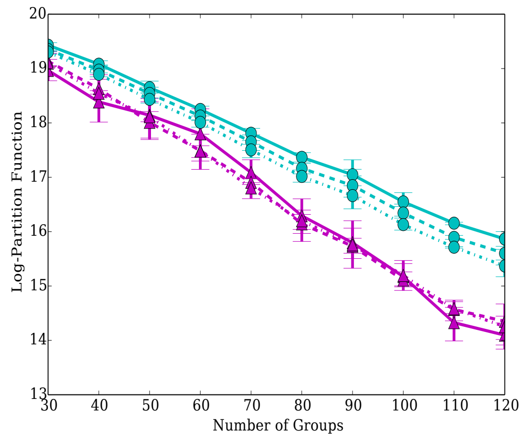
Since $\widehat{\mathcal{Z}}_L(F_{\{i\}}) = \max_{X\in V\setminus\{i\}} e^{F(\{i\})-F(X\cup\{i\})} \prod_{j\in X}(1+ e^{F(V)-F(V\setminus\{j\})}) \prod_{j\notin X}(1 + e^{-F(\{i,j\})+F(\{i\})})$ and $\widehat{\mathcal{Z}}_L(F^{V\setminus\{i\}}) = \max_{X\in V\setminus\{i\}} e^{-F(X)} \prod_{j\in X}(1 + e^{F(V\setminus\{i\})-F(V\setminus\{i,j\})}) \prod_{j\notin X}(1+e^{-F(\{j\})})$, we explicitly write the lower bound after clamping as follows.

$$\begin{aligned}
& e^{-F(\{i\})}\widehat{\mathcal{Z}}_L(F_{\{i\}}) + \widehat{\mathcal{Z}}_L(F^{V\setminus\{i\}}) \\
& = \max_{X\in V\setminus\{i\}} e^{-F(X\cup\{i\})} \prod_{j\in X} (1+e^{F(V)-F(V\setminus\{j\})}) \cdot \\
& \prod_{j\notin X} (1+e^{F(\{i\})-F(\{i,j\})}) + \max_{X\in V\setminus\{i\}} e^{-F(X)} \cdot \\
& \prod_{j\in X} (1+e^{F(V\setminus\{i\})-F(V\setminus\{i,j\})}) \prod_{j\notin X} (1+e^{-F(\{j\})}) \\
& := A + B
\end{aligned} \tag{14}$$

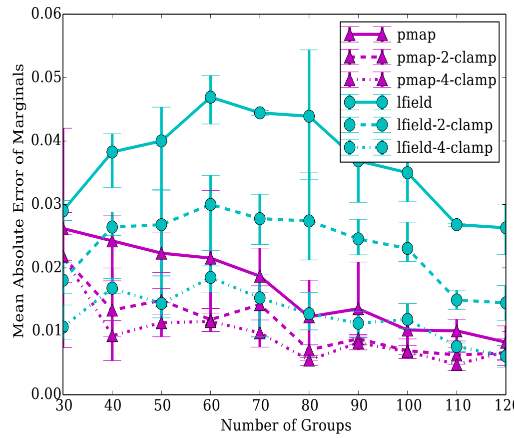
We claim that if $i \in X^*$, $A \geq A_1, B \geq B_1$, hence $A + B \geq A_1 + B_1$, and if $i \notin X^*$, $B \geq A_2, A \geq B_2$, hence $A+B \geq A_2+B_2$. If this is true, then the expected result follows.

Let $X = X^* \setminus \{i\}$ when $i \in X^*$, then $A_1 = e^{-F(X\cup\{i\})} \prod_{j\in X}(1+e^{F(V)-F(V\setminus\{j\})}) \prod_{j\notin X\cup\{i\}}(1+ e^{-F(\{j\})})$. We compare A_1 with $A = e^{-F(X\cup\{i\})} \prod_{j\in X}(1 + e^{F(V)-F(V\setminus\{j\})}) \prod_{j\notin X}(1 + e^{F(\{i\})-F(\{i,j\})})$. Since $F(\{i\}) - F(\{i,j\}) \geq -F(\{j\})$ by diminishing return, it is easy to see $A \geq A_1$. On the other hand, $B_1 = e^{F(V)-F(V\setminus\{i\})-F(X\cup\{i\})} \prod_{j\in X}(1 + e^{F(V)-F(V\setminus\{j\})}) \prod_{j\notin X\cup\{i\}}(1 + e^{-F(\{j\})})$. We compare this with $B = e^{-F(X)} \prod_{j\in X}(1 + e^{F(V\setminus\{i\})-F(V\setminus\{i,j\})}) \prod_{j\notin X}(1 + e^{-F(\{j\})})$. Since $F(V) - F(V \setminus \{i\}) - F(X \cup \{i\}) \leq -F(X)$ and $F(V) - F(V \setminus \{j\}) \leq F(V \setminus \{i\}) - F(V \setminus \{i,j\})$ by diminishing return, it follows that $B \geq B_1$.

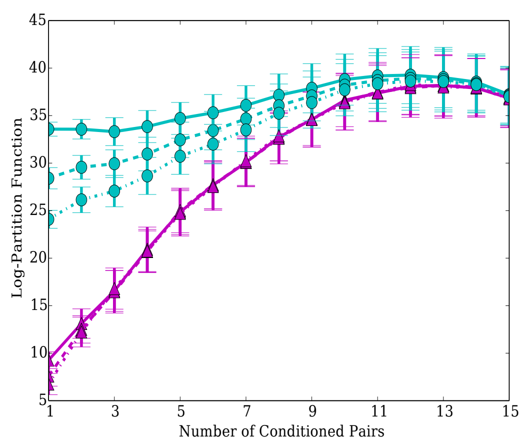
Let $X = X^*$ when $i \notin X^*$, then $B_2 = e^{-F(X)-F(\{i\})} \prod_{j\in X}(1 + e^{F(V)-F(V\setminus\{j\})}) \prod_{j\notin X\cup\{i\}}(1+e^{-F(\{j\})})$. We compare this with A. Since $-F(X) - F(\{i\}) \leq -F(X \cup \{i\})$ and $-F(\{j\}) \leq F(\{i\}) - F(\{i,j\})$, $A \geq B_2$ follows. Moreover, $A_2 = e^{-F(X)} \prod_{j\in X}(1 + e^{F(V)-F(V\setminus\{j\})}) \prod_{j\notin X\cup\{i\}}(1 + e^{-F(\{j\})})$, hence $B \geq A_2$ follows because $F(V) - F(V \setminus \{j\}) \leq F(V \setminus \{i\}) - F(V \setminus \{i,j\})$. This completes the proof. □



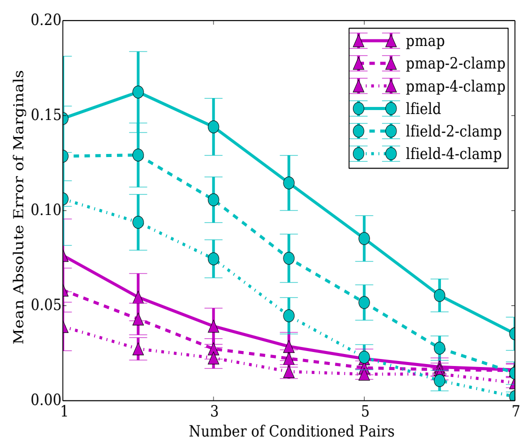
(a) random covers ($m = 30, \alpha = 0.5, \beta = \frac{1}{30}$)



(b) random covers ($m = 10, \alpha = 0.5, \beta = \frac{1}{10}$)



(c) conditioned pairs ($n = 50, c = 1$)



(d) conditioned pairs ($n = 20, c = 1$)

Figure 4: Additional experiments on random covers and conditioned pairs with different parameters. Still we can see that clamping improves the estimates on both $\mathcal{Z}$ and the marginals.