7 Appendix

7.1 Variational forms of convex envelopes (Proof of lemma 2 and Remark 1)

In this section, we recall the different variational forms of the homogeneous convex envelope derived in [31] and derive similar variational forms for the non-homogeneous convex envelope, which includes the ones stated in lemma 2). These variational forms will be needed in some of our proofs below.

Lemma 4. *The homogeneous convex envelope Ω_p of F_p admits the following variational forms.*

$$\Omega_\infty(w) = \min_\alpha \{ \sum_{S \subseteq V} \alpha_S F(S) : \sum_{S \subseteq V} \alpha_S \mathbb{1}_S \geq |w|, \alpha_S \geq 0 \}. \tag{9}$$

$$\Omega_p(w) = \min_v \{ \sum_{S \subseteq V} F(S)^{1/q} \|v^S\|_p : \sum_{S \subseteq V} v^S = |w|, \mathrm{supp}(v^S) \subseteq S \}. \tag{10}$$

$$= \max_{\kappa \in \mathbb{R}^d_+} \sum_{i=1}^d \kappa_i^{1/q} |w_i| \textit{ s.t. } \kappa(A) \leq F(A), \forall A \subseteq V. \tag{11}$$

$$= \inf_{\eta \in \mathbb{R}^d_+} \frac{1}{p} \sum_{j=1}^d \frac{|w_j|^p}{\eta_j^{p-1}} + \frac{1}{q} \Omega_\infty(\eta). \tag{12}$$

The non-homogeneous convex envelope of a set function F, over the unit ℓ_∞-ball was derived in [10], where it was shown that $\Theta_\infty(w) = \inf_{\eta \in [0,1]^d} \{ f(\eta) : \eta \geq |w| \}$ where f is any proper, l.s.c. convex *extension* of F (c.f., Lemma 1 [10]). A natural choice for f is the *convex closure* of F, which corresponds to the *tightest* convex extension of F on $[0,1]^d$. We recall the two equivalent definitions of convex closure, which we have adjusted to allow for infinite values.

Definition 5 (Convex Closure; c.f., [9, Def. 3.1]). *Given a set function $F : 2^V \to \overline{\mathbb{R}}$, the convex closure $f^- : [0,1]^d \to \overline{\mathbb{R}}$ is the point-wise largest convex function from $[0,1]^d$ to $\overline{\mathbb{R}}$ that always lowerbounds F.*

Definition 6 (Equivalent definition of Convex Closure; c.f., [35, Def. 1] and [9, Def. 3.2]). *Given any set function $f : \{0,1\}^n \to \overline{\mathbb{R}}$, the convex closure of f can equivalently be defined $\forall w \in [0,1]^n$ as:*

$$f^-(w) = \inf \{ \sum_{S \subseteq V} \alpha_S F(S) : w = \sum_{S \subseteq V} \alpha_S \mathbb{1}_S, \sum_{S \subseteq V} \alpha_S = 1, \alpha_S \geq 0 \}$$

It is interesting to note that $f^-(w) = f_L(w)$ where f_L is Lovász extension iff F is a submodular function [35].

The following lemma derive variational forms of Θ_p for any $p \geq 1$ that parallel the ones known for the homogeneous envelope.

Lemma 5. *The non-homogeneous convex envelope Θ_p of F_p admits the following variational forms.*

$$\Theta_\infty(w) = \inf \{ \sum_{S \subseteq V} \alpha_S F(S) : \sum_{S \subseteq V} \alpha_S \mathbb{1}_S \geq |w|, \sum_{S \subseteq V} \alpha_S = 1, \alpha_S \geq 0 \}. \tag{13}$$

$$\Theta_p(w) = \max_{\kappa \in \mathbb{R}^d} \sum_{j=1}^d \psi_j(\kappa_j, w_j) + \min_{S \subseteq V} F(S) - \kappa(S), \ \forall w \in \mathrm{dom}(\Theta_p(w)). \tag{14}$$

$$= \inf_{\eta \in [0,1]^d} \frac{1}{p} \sum_{j=1}^d \frac{|w_j|^p}{\eta_j^{p-1}} + \frac{1}{q} f^-(\eta), \tag{15}$$

where $\mathrm{dom}(\Theta_p) = \{ w | \exists \eta \in [0,1]^d \textit{ s.t } \mathrm{supp}(w) \subseteq \mathrm{supp}(\eta), \eta \in \mathrm{dom}(f^-) \}$*, and where we define*

$$\psi_j(\kappa_j, w_j) := \begin{cases} \kappa_j^{1/q} |w_j| & \textit{if } |w_j| \leq \kappa_j^{1/p}, \kappa_j \geq 0 \\ \frac{1}{p} |w_j|^p + \frac{1}{q} \kappa_j & \textit{otherwise.} \end{cases}$$

If F is monotone, $\Theta_\infty = f^-$, then we can replace f^- by Θ_∞ in (15) *and we can restrict $\kappa \in \mathbb{R}^d_+$ in* (14).

To prove the variational form (13) in Lemma 5, we need to show first the following property of f^-.

Proposition 5 (c.f., [9, Prop. 3.23]). *The minimum values of a proper set function F and its convex closure f^- are equal, i.e.,*

$$\min_{w\in[0,1]^d} f^-(w) = \min_{S\subseteq V} F(S)$$

If S is a minimizer of $f(S)$, then $\mathbb{1}_S$ is a minimizer of f^-. Moreover, if w is a minimizer of f^-, then every set in the support of α, where $f^-(w) = \sum_{S\subseteq V} \alpha_S F(S)$, is a minimizer of F.

Proof. First note that, $\{0,1\}^d \subseteq [0,1]^d$ implies that $f^-(w^*) \leq F(S^*)$. On the other hand, $f^-(w^*) = \sum_{S\subseteq V} \alpha^*_S F(S) \geq \sum_{S\subseteq V} \alpha^*_S F(S^*) = F(S^*)$. The rest of the proposition follows directly. □

Given the choice of the extension $f = f^-$, the variational form (13) of Θ_∞ given in lemma 5 follows directly from definition 6 and proposition 5, as shown in the following corollary.

Corollary 4. *Given any set function $F : 2^V \to \overline{\mathbb{R}}_+$ and its corresponding convex closure f^-, the convex envelope of $F(\mathrm{supp}(w))$ over the unit ℓ_∞-ball is given by*

$$\begin{aligned}\Theta_\infty(w) &= \inf_\alpha\{\sum_{S\subseteq V} \alpha_S F(S) : \sum_{S\subseteq V} \alpha_S \mathbb{1}_S \geq |w|, \sum_{S\subseteq V} \alpha_S = 1, \alpha_S \geq 0\}.\\ &= \inf_v\{\sum_{S\subseteq V} F(S)\|v^S\|_\infty : \sum_{S\subseteq V} v^S = |w|, \sum_{S\subseteq V} \|v^S\|_\infty = 1, \mathrm{supp}(v^S) \subseteq S\}.\end{aligned}$$

Proof. f^- satisfies the first 2 assumptions required in Lemma 1 of [10], namely, f^- is a lower semi-continuous convex extension of F which satisfies

$$\max_{S\subseteq V} m(S) - F(S) = \max_{w\in[0,1]^d} m^T w - f^-(w), \forall m \in \mathbb{R}^d_+$$

To see this note that $m^T w^* - f^-(w^*) = \sum_{S\subseteq V} \alpha^*_S (m^T \mathbb{1}_S - F(S)) \geq \sum_{S\subseteq V} \alpha^*_S (m^T \mathbb{1}_{S^*} - F(S^*)) = m(S^*) - F(S^*)$. The other inequality is trivial. The corollary then follows directly from Lemma 1 in [10] and definition 6. □

Note that $\mathrm{dom}(\Theta_\infty) = \{w : \exists \eta \in [0,1]^d \cap \mathrm{dom}(f^-), \eta \geq |w|\}$. Note also that Θ_∞ is monotone even if F is not. On the other hand, if F is monotone, then f^- is monotone on $[0,1]^d$ and $\Theta_\infty(w) = f^-(|w|)$. Then the proof of remark 1 follows, since if F is a monotone submodular function and f_L is its Lovász extension, then $\Theta_\infty(w) = f^-(|w|) = f_L(|w|) = \Omega_\infty(w), \forall w \in [-1,1]^d$, where the last equality was shown in [1].

Next, we derive the convex relaxation of F_p for a general $p \geq 1$.

Proposition 6. *Given any set function $F : 2^V \to \overline{\mathbb{R}}_+$ and its corresponding convex closure f^-, the convex envelope of $F_{\mu\lambda}(w) = \mu F(\mathrm{supp}(w)) + \lambda\|w\|^p_p$ is given by*

$$\Theta_p(w) = \inf_{\eta\in[0,1]^d} \lambda \sum_{j=1}^d \frac{|w_j|^p}{\eta_j^{p-1}} + \mu f^-(\eta).$$

Note that $\mathrm{dom}(\Theta_p) = \{w | \exists \eta \in [0,1]^d$ *s.t* $\mathrm{supp}(w) \subseteq \mathrm{supp}(\eta), \eta \in \mathrm{dom}(f^-)\}$.

Proof. Given any proper l.s.c. convex extension f of F, we have:

First for the case where $p = 1$:

$$\begin{aligned}F^*_{\mu\lambda}(s) &= \sup_{w\in\mathbb{R}^n} w^T s - \mu F(\mathrm{supp}(w)) - \lambda\|w\|_1\\ &= \sup_{\eta\in\{0,1\}^d} \sup_{\substack{\mathbb{1}_{\mathrm{supp}(w)}=\eta\\ \mathrm{sign}(w)=\mathrm{sign}(s)}} |w|^T(|s| - \lambda\mathbb{1}) - \mu F(\eta)\\ &= \iota_{\{|s|\leq\lambda\mathbb{1}\}}(s) - \inf_{\eta\in\{0,1\}^d} \mu F(\eta).\end{aligned}$$

Hence $F^{**}_{\mu\lambda}(w) = \lambda\|w\|_1 + \inf_{\eta\in\{0,1\}^d} \lambda F(\eta)$. For the case $p \in (1, \infty)$.

$$\begin{aligned}
F^{*}_{\mu\lambda}(s) &= \sup_{w\in\mathbb{R}^d} w^T s - \mu F(\mathrm{supp}(w)) - \lambda\|w\|_p^p \\
&= \sup_{\eta\in\{0,1\}^d} \sup_{\substack{\mathbb{1}_{\mathrm{supp}(w)}=\eta \\ \mathrm{sign}(w)=\mathrm{sign}(s)}} |w|^T|s| - \lambda\|w\|_p^p - \mu F(\eta) \\
&= \sup_{\eta\in\{0,1\}^d} \frac{\lambda(p-1)}{(\lambda p)^q}\eta^T|s|^q - \mu F(\eta) && (|s_i| = \lambda p|x_i^*|^{p-1}, \forall \eta_i \neq 0) \\
&= \sup_{\eta\in[0,1]^d} \frac{\lambda(p-1)}{(\lambda p)^q}\eta^T|s|^q - \mu f^-(\eta).
\end{aligned}$$

We denote $\hat\lambda = \frac{\lambda(p-1)}{(\lambda p)^q}$.

$$\begin{aligned}
F^{**}_{\mu\lambda}(w) &= \sup_{s\in\mathbb{R}^d} w^T s - F^{*}_{\mu\lambda}(s) \\
&= \sup_{s\in\mathbb{R}^d} \min_{\eta\in[0,1]^d} s^T w - \hat\lambda\eta^T|s|^q + \mu f^-(\eta) \\
&\overset{\star}{=} \inf_{\eta\in[0,1]^d} \sup_{\substack{s\in\mathbb{R}^p \\ \mathrm{sign}(s)=\mathrm{sign}(w)}} |s|^T|w| - \hat\lambda\eta^T|s|^q + \mu f^-(\eta) \\
&= \inf_{\eta\in[0,1]^d} \lambda(|w|^p)^T\eta^{1-p} + \mu f^-(\eta),
\end{aligned}$$

where the last equality holds since $|w_i| = \hat\lambda\eta_i q|s_i^*|^{q-1}, \forall\eta_i \neq 0$, otherwise $s_i^* = 0$ if $w_i = 0$ and ∞ otherwise. ($\star$) holds by Sion's minimax theorem [34, Corollary 3.3]. Note then that the minimizer η^* (if it exists) satisfies $\mathrm{supp}(w) \subseteq \mathrm{supp}(\eta^*)$. Finally, note that if we take the limit as $p \to \infty$, we recover $\Theta_\infty = \inf_{\eta\in[0,1]^d}\{f^-(\eta) : \eta \geq |x|\}$. □

The variational form (15) given in lemma 5 follows from proposition 6 for the choice $\mu = \frac{1}{q}, \lambda = \frac{1}{p}$.

The following proposition derives the variational form (14) for $p = \infty$.

Proposition 7. *Given any set function $F : 2^V \to \mathbb{R}\cup\{+\infty\}$, and its corresponding convex closure f^-, Θ_∞ can be written $\forall w \in \mathrm{dom}(\Theta_\infty)$ as*

$$\begin{aligned}
\Theta_\infty(w) &= \max_{\kappa\in\mathbb{R}^d_+}\{\kappa^T|w| + \min_{S\subseteq V} F(S) - \kappa(S)\} \\
&= \max_{\kappa\in\mathbb{R}^d_+}\{\kappa^T|w| + \min_{S\subseteq\mathrm{supp}(w)} F(S) - \kappa(S)\} && \text{(if } F \text{ is monotone)}
\end{aligned}$$

Similarly $\forall w \in \mathrm{dom}(f^-)$ we can write

$$\begin{aligned}
f^-(w) &= \max_{\kappa\in\mathbb{R}^d}\{\kappa^T|w| + \min_{S\subseteq V} F(S) - \kappa(S)\} \\
&= \Theta_\infty(w) = \max_{\kappa\in\mathbb{R}^d_+}\{\kappa^T w + \min_{S\subseteq\mathrm{supp}(w)} F(S) - \kappa(S)\} && \text{(if } F \text{ is monotone)}
\end{aligned}$$

Proof. $\forall w \in \mathrm{dom}(\Theta_\infty)$, strong duality holds by Slater's condition, hence

$$\begin{aligned}
\Theta_\infty(w) &= \min_{\alpha}\{\sum_{S\subseteq V}\alpha_S F(S) : \sum_{S\subseteq V}\alpha_S\mathbb{1}_S \geq |w|, \sum_{S\subseteq V}\alpha_S = 1, \alpha_S \geq 0\}. \\
&= \min_{\alpha\geq 0} \max_{\rho\in\mathbb{R},\kappa\in\mathbb{R}^d_+}\{\sum_{S\subseteq V}\alpha_S F(S) + \kappa^T(|w| - \sum_{S\subseteq V}\alpha_S\mathbb{1}_S) + \rho(1 - \sum_{S\subseteq V}\alpha_S)\}. \\
&= \max_{\rho\in\mathbb{R},\kappa\in\mathbb{R}^d_+} \min_{\alpha\geq 0}\{\kappa^T|w| + \sum_{S\subseteq V}\alpha_S(F(S) - \kappa^T\mathbb{1}_S - \rho) + \rho\}. \\
&= \max_{\rho\in\mathbb{R},\kappa\in\mathbb{R}^d_+}\{\kappa^T|w| + \rho : F(S) \geq \kappa^T\mathbb{1}_S + \rho)\}. \\
&= \max_{\kappa\in\mathbb{R}^d_+}\{\kappa^T|w| + \min_{S\subseteq V} F(S) - \kappa(S)\}.
\end{aligned}$$

Let $J = \text{supp}(|w|)$ then $\kappa^*_{J^c} = 0$. Then for monotone functions $F(S) - \kappa^*(S) \geq F(S \cap J) - \kappa^*(S)$, so we can restrict the minimum to $S \subseteq J$. The same proof holds for f^-, with the Lagrange multiplier $\kappa \in \mathbb{R}^d$ not constrained to be positive. □

The following Corollary derives the variational form (14) for $p \in [1, \infty]$.

Corollary 5. *Given any set function $F : 2^V \to \mathbb{R} \cup \{+\infty\}$, Θ_p can be written $\forall w \in \text{dom}(\Theta_p)$ as*

$$\Theta_p(w) = \max_{\kappa \in \mathbb{R}^d} \sum_{j=1}^d \psi_j(\kappa_j, w_j) + \min_{S \subseteq V} F(S) - \kappa(S).$$

$$= \max_{\kappa \in \mathbb{R}^d_+} \sum_{j=1}^d \psi_j(\kappa_j, w_j) + \min_{S \subseteq V} F(S) - \kappa(S). \qquad \text{(if } F \text{ is monotone)}$$

where

$$\psi_j(\kappa_j, w_j) := \begin{cases} \kappa_j^{1/q}|w_j| & \textit{if } |w_j| \leq \kappa_j^{1/p}, \kappa_j \geq 0 \\ \frac{1}{p}|w_j|^p + \frac{1}{q}\kappa_j & \textit{otherwise} \end{cases}$$

Proof. By Propositions 6 and 7, we have $\forall w \in \text{dom}(\Theta_p)$, i.e., $\exists \eta \in [0,1]^d$, s.t $\text{supp}(w) \subseteq \text{supp}(\eta), \eta \in \text{dom}(f^-)$,

$$\Theta_p(w) = \inf_{\eta \in [0,1]^d} \frac{1}{p} \sum_{j=1}^d \frac{|w_j|^p}{\eta_j^{p-1}} + \frac{1}{q} f^-(\eta)$$

$$= \inf_{\eta \in [0,1]^d} \frac{1}{p} \sum_{j=1}^d \frac{|w_j|^p}{\eta_j^{p-1}} + \frac{1}{q} \max_{\rho \in \mathbb{R}, \kappa \in \mathbb{R}^d} \{\kappa^T \eta + \rho : F(S) \geq \kappa^T \mathbb{1}_S + \rho\}.$$

$$\overset{\star}{=} \max_{\rho \in \mathbb{R}, \kappa \in \mathbb{R}^d} \inf_{\eta \in [0,1]^d} \{\frac{1}{p} \sum_{j=1}^d \frac{|w_j|^p}{\eta_j^{p-1}} + \frac{1}{q} \kappa^T \eta + \rho : F(S) \geq \kappa^T \mathbb{1}_S + \rho\}.$$

($\star$) holds by Sion's minimax theorem [34, Corollary 3.3]. Note also that

$$\inf_{\eta_j \in [0,1]} \frac{1}{p} \frac{|w_j|^p}{\eta_j^{p-1}} + \frac{1}{q} \kappa_j \eta_j = \begin{cases} \kappa_j^{1/q}|w_j| & \text{if } |w_j| \leq \kappa_j^{1/p}, \kappa_j \geq 0 \\ \frac{1}{p}|w_j|^p + \frac{1}{q}\kappa_j & \text{otherwise} \end{cases} := \psi_j(\kappa_j, w_j)$$

where the minimum is $\eta_j^* = 1$ if $\kappa_j \leq 0$. If $\kappa_j \geq 0$, the infimum is zero if $w_j = 0$. Otherwise, the minimum is achieved at $\eta_j^* = \min\{\frac{|w_j|}{\kappa_j^{1/p}}, 1\}$ (if $\kappa_j = 0, \eta_j^* = 1$). Hence,

$$\Theta_p(w) = \max_{\kappa \in \mathbb{R}^d} \sum_{j=1}^d \psi_j(\kappa_j, w_j) + \min_{S \subseteq V} F(S) - \kappa(S).$$

□

7.2 Necessary conditions for support recovery (Proof of Theorem 1)

Before proving Theorem 1, we need the following technical Lemma.

Lemma 6. *Given $J \subset V$ and a vector w s.t $\text{supp}(w) \subseteq J$, if Φ is not decomposable at w w.r.t J, then $\exists i \in J^c$ such that the i-th component of all subgradients at w is zero; $0 = [\partial\Phi(w)]_i$.*

Proof. If Φ is not decomposable at w and $0 \neq [\partial\Phi(w)]_i, \forall i \in J^c$, then $\forall M_J > 0, \exists \Delta \neq 0, \text{supp}(\Delta) \subseteq J^c$ s.t., $\Phi(w + \Delta) < \Phi(w) + M_J\|\Delta\|_\infty$. In particular, we can choose $M_J = \inf_{i \in J^c, v \in \partial\Phi(w_J), v_i \neq 0} |v_i| > 0$, if the inequality holds for some $\Delta \neq 0$, then let $i_{\max}$ denote the index where $|\Delta_{i_{\max}}| = \|\Delta\|_\infty$. Then given any $v \in \partial\Phi(w)$ s.t., $v_{i_{\max}} \neq 0$, we have

$$\begin{aligned} \Phi(w + \|\Delta\|_\infty \mathbb{1}_{i_{\max}}) \leq \Phi(w + \Delta) &< \Phi(w) + M_J\|\Delta\|_\infty \\ &\leq \Phi(w) + \langle v, \|\Delta\|_\infty \mathbb{1}_{i_{\max}} \text{sign}(v_{i_{\max}})\rangle \\ &\leq \Phi(w + \|\Delta\|_\infty \mathbb{1}_{i_{\max}}) \end{aligned}$$

which leads to a contradiction. □

Theorem 1. *The minimizer $\hat{w}$ of $\min_{w\in\mathbb{R}^d} L(w) - z^\top w + \lambda\Phi(w)$, where L is a strongly-convex and smooth loss function and $z \in \mathbb{R}^d$ has a continuous density w.r.t to the Lebesgue measure, has a weakly stable support w.r.t. Φ, with probability one.*

Proof. We will show in particular that Φ is decomposable at $\hat{w}$ w.r.t $\text{supp}(\hat{w})$. Since L is strongly-convex, given z the corresponding minimizer $\hat{w}$ is unique, then the function $h(z) := \arg\min_{w\in\mathbb{R}^d} L(w) - z^T w + \lambda\Phi(w)$ is well defined. We want to show that

$$\begin{aligned}
&P(\forall z,\ \Phi \text{ is decomposable at } h(z) \text{ w.r.t } \text{supp}(h(z))\ )\\
&= 1 - P(\exists z, \text{s.t, } \Phi \text{ is not decomposable at } h(z) \text{ w.r.t } \text{supp}(h(z))\ )\\
&\geq 1 - P(\exists z,\ \text{s.t., } \exists i \in (\text{supp}(h(z)))^c, [\partial\Phi(h(z))]_i = 0) \qquad \text{by lemma 6}\\
&= 1.
\end{aligned}$$

Given fixed $i \in V$, we show that the set $S_i := \{z : i \in (\text{supp}(h(z)))^c, [\partial\Phi(h(z))]_i = 0\}$ has measure zero. Then, taking the union of the finitely many sets $S_i, \forall i \in V$, all of zero measure, we have $P(\exists z,\ \text{s.t., } \exists i \in (\text{supp}(h(z)))^c, [\partial\Phi(h(z))]_i = 0) = 0$.

To show that the set S_i has measure zero, let $z_1, z_2 \in S_i$ and denote by $\mu > 0$ the strong convexity constant of L. We have by convexity of Φ:

$$\begin{aligned}
\Big(\big(z_1 - \nabla L(h(z_1))\big) - \big(z_2 - \nabla L(h(z_2))\big)\Big)^\top \Big(h(z_1) - h(z_2)\Big) &\geq 0\\
(z_1 - z_2)^\top (h(z_1) - h(z_2)) &\geq \big(\nabla L(h(z_1)) - \nabla L(h(z_2))\big)^\top \big(h(z_1) - h(z_2)\big)\\
(z_1 - z_2)^\top (h(z_1) - h(z_2)) &\geq \mu\|h(z_1) - h(z_2)\|_2^2\\
\frac{1}{\mu}\|z_1 - z_2\|_2 &\geq \|h(z_1) - h(z_2)\|_2
\end{aligned}$$

Thus h is a deterministic Lipschitz-continuous function of z. Let $J = \text{supp}(h(z))$, then by optimality conditions $z_J - \nabla L(h(z_J)) \in \partial\Phi(h(z_J))$ (since $h(z) = h(z_J)$), then $z_i - \nabla L(h(z_J))_i = 0$ since $[\partial\Phi(h(z_J))]_i = 0$. and thus z_i is a Lipschitz-continuous function of z_J, which can only happen with zero measure. □

7.3 Sufficient conditions for support recovery (Proof of Lemma 3 and Theorem 2)

Lemma 3. *Let Φ be a monotone convex function, $\Phi(|w|^\alpha)$ admits the following majorizer, $\forall w^0 \in \mathbb{R}^d$, $\Phi(|w|^\alpha) \leq (1-\alpha)\Phi(|w^0|^\alpha) + \alpha\Phi(|w^0|^{\alpha-1} \circ |w|)$, which is tight at w^0.*

Proof. The function $w \to w^\alpha$ is concave on $\mathbb{R}_+ \setminus \{0\}$, hence

$$\begin{aligned}
|w_j|^\alpha &\leq |w_j^0|^\alpha + \alpha|w_j^0|^{\alpha-1}(|w_j| - |w_j|^0)\\
|w_j|^\alpha &\leq (1-\alpha)|w_j^0|^\alpha + \alpha|w_j^0|^{\alpha-1}|w_j|\\
\Phi(|w|^\alpha) &\leq \Phi((1-\alpha)|w^0|^\alpha + \alpha|w^0|^{\alpha-1} \circ |w_j|) && \text{(by monotonicity)}\\
\Phi(|w|^\alpha) &\leq (1-\alpha)\Phi(|w^0|^\alpha) + \alpha\Phi(|w^0|^{\alpha-1} \circ |w|) && \text{(by convexity)}
\end{aligned}$$

If $w_j = 0$ for any j, the upper bound goes to infinity and hence it still holds. □

Theorem 2. *[Consistency and Support Recovery] Let $\Phi : \mathbb{R}^d \to \overline{\mathbb{R}}_+$ be a proper normalized absolute monotone convex function and denote by J the true support $J = \text{supp}(w^*)$. If $|w^*|^\alpha \in \text{int dom}\,\Phi$, J is strongly stable with respect to Φ and λ_n satisfies $\frac{\lambda_n}{\sqrt{n}} \to 0$, $\frac{\lambda_n}{n^{\alpha/2}} \to \infty$, then the estimator (6) is consistent and asymptotically normal, i.e., it satisfies*

$$\sqrt{n}(\hat{w}_J - w_J^*) \xrightarrow{d} \mathcal{N}(0, \sigma^2 Q_{JJ}^{-1}), \tag{7}$$

and

$$P(\text{supp}(\hat{w}) = J) \to 1. \tag{8}$$

Proof. We will follow the proof in [38]. We write $\hat{w} = w^* + \frac{\hat{u}}{\sqrt{n}}$ and $\Psi_n(u) = \frac{1}{2}\|y - X(w^* + \frac{u}{\sqrt{n}})\|_2^2 + \lambda_n \Phi(c \circ |w^* + \frac{u}{\sqrt{n}}|)$, where $c = |w^0|^{\alpha-1}$. Then $\hat{u} = \arg\min_{u \in \mathbb{R}^d} \Psi_n(u)$. Let $V_n(u) = \Psi_n(u) - \Psi_n(0)$, then

$$V_n(u) = \frac{1}{2} u^T Q u - \epsilon^T \frac{Xu}{\sqrt{n}} + \lambda_n \big(\Phi(c \circ |w^* + \frac{u}{\sqrt{n}}|) - \Phi(c \circ |w^*|)\big)$$

Since w^0 is a $\sqrt{n}$-consistent estimator to w^*, then $\sqrt{n} w^0_{J^c} = O_p(1)$ and $n^{\frac{1-\alpha}{2}} c_{J^c}^{-1} = O_p(1)$. Since $\frac{\lambda_n}{n^{\alpha/2}} \to \infty$, by stability of J, we have

$$\begin{aligned}
\lambda_n \big(\Phi(c \circ |w^* + \frac{u}{\sqrt{n}}|) - \Phi(c \circ |w^*|)\big) &= \lambda_n \big(\Phi(c_J \circ |w_J^* + \frac{u_J}{\sqrt{n}}| + c_{J^c} \circ \frac{|u_{J^c}|}{\sqrt{n}}) - \Phi(c_J \circ |w_J^*|)\big) \\
&\geq \lambda_n \big(\Phi(c_J \circ |w_J^* + \frac{u_J}{\sqrt{n}}|) + M_J \|c_{J^c} \circ \frac{|u_{J^c}|}{\sqrt{n}}\|_\infty - \Phi(c_J \circ |w_J^*|)\big) \\
&= \lambda_n \big(\Phi(c_J \circ |w_J^* + \frac{u_J}{\sqrt{n}}|) - \Phi(c_J \circ |w_J^*|)\big) + M_J \|\lambda_n n^{-\alpha/2} n^{\frac{\alpha-1}{2}} c_{J^c} \circ |u_{J^c}|\|_\infty \\
&\xrightarrow{p} \infty \quad \text{if } u_{J^c} \neq 0
\end{aligned} \tag{16}$$

Otherwise, if $u_{J^c} = 0$, we argue that

$$\lambda_n \big(\Phi(c \circ |w^* + \frac{u}{\sqrt{n}}|) - \Phi(c \circ |w^*|)\big) = \lambda_n (\Phi(c_J \circ |w_J^* + \frac{u_J}{\sqrt{n}}|) - \Phi(c_J \circ |w_J^*|)) \xrightarrow{p} 0. \tag{17}$$

To see this note first that since w^0 is a $\sqrt{n}$-consistent estimator to w^*, then $c_J = |w_J^0|^{\alpha-1} \xrightarrow{p} |w_J^*|^{\alpha-1}$, $c_J \circ |w_J^*| \xrightarrow{p} |w_J^*|^\alpha$ and $c_J \circ |w_J^* + \frac{u_J}{\sqrt{n}}| \xrightarrow{p} |w_J^*|^\alpha$. Then by the assumption $|w^*|^\alpha \in \text{int dom } \Phi$, we have that both $c_J \circ |w_J^*|, c_J \circ |w_J^* + \frac{u_J}{\sqrt{n}}| \in \text{int dom } \Phi$ with probability going to one. By convexity, we then have:

$$\begin{aligned}
\lambda_n (\Phi(c_J \circ |w_J^* + \frac{u_J}{\sqrt{n}}|) - \Phi(c_J \circ |w_J^*|)) &\geq \langle \nabla \Phi(c_J \circ |w_J^*|), \lambda_n \frac{u_J}{\sqrt{n}} \rangle \\
\lambda_n (\Phi(c_J \circ |w_J^* + \frac{u_J}{\sqrt{n}}|) - \Phi(c_J \circ |w_J^*|)) &\leq \langle \nabla \Phi(c_J \circ |w_J^* + \frac{u_J}{\sqrt{n}}|), \lambda_n \frac{u_J}{\sqrt{n}} \rangle
\end{aligned}$$

where $\nabla \Phi(w)$ denotes a subgradient of Φ at w.

For all $w \in \text{int dom } \Phi$ where Φ is convex, monotone and normalized, we have that $\|z\|_\infty < \infty, \forall z \in \partial \Phi(w)$. To see this, note that since $w \in \text{int dom } \Phi$, $\exists \delta > 0$ s.t., $\forall w' \in B_\delta(w), \Phi(w') < +\infty$. Let $w' = w + \text{sign}(z) \mathbb{1}_{i_{\max}} \delta$, where $i_{\max}$ denotes the index where $|z_{i_{\max}}| = \|z\|_\infty$ then by convexity we have

$$\begin{aligned}
\Phi(w') &\geq \Phi(w) + \langle z, w' - w \rangle, & \forall z \in \partial \Phi(w) \\
+\infty > \Phi(w') &\geq \|z\|_\infty \delta, & \forall z \in \partial \Phi(w), \quad (\text{since } \Phi(w) \geq 0)
\end{aligned}$$

Since $\frac{\lambda_n}{\sqrt{n}} \to 0$, we can then conclude by Slutsky's theorem that (17) holds.

Hence by (16) and (17),

$$\lambda_n \big(\Phi(c \circ |w^* + \frac{u}{\sqrt{n}}|) - \Phi(c \circ |w^*|)\big) \xrightarrow{p} \begin{cases} 0 & \text{if } u_{J^c} = 0 \\ \infty & \text{Otherwise} \end{cases}. \tag{18}$$

By CLT, $\frac{X^T \epsilon}{\sqrt{n}} \xrightarrow{d} W \sim \mathcal{N}(0, \sigma^2 Q)$, it follows then that $V_n(u) \xrightarrow{d} V(u)$, where

$$V(u) = \begin{cases} \frac{1}{2} u_J^T Q_{JJ} u_J - W_J^T u_J & \text{if } u_{J^c} = 0 \\ \infty & \text{Otherwise} \end{cases}.$$

V_n is convex and the unique minimum of V is $u_J = Q_{JJ}^{-1} W_J, u_{J^c} = 0$, hence by epi-convergence results [c.f., [38]]

$$\hat{u}_J \xrightarrow{d} Q_{JJ}^{-1} W_J \sim \mathcal{N}(0, \sigma^2 Q_{JJ}^{-1}), \quad \hat{u}_{J^c} \xrightarrow{d} 0. \tag{19}$$

Since $\hat{u} = \sqrt{n}(\hat{w} - w^*)$, then it follows from (19) that

$$\hat{w}_J \xrightarrow{p} w_J^*, \quad \hat{w}_{J^c} \xrightarrow{p} 0 \tag{20}$$

Hence, $P(\mathrm{supp}(\hat{w}) \supseteq J) \to 1$ and it is sufficient to show that $P(\mathrm{supp}(\hat{w}) \subseteq J) \to 1$ to complete the proof.

For that denote $\hat{J} = \mathrm{supp}(\hat{w})$ and let's consider the event $\hat{J} \setminus J \neq \emptyset$. By optimality conditions, we know that

$$-X_{\hat{J}\setminus J}^T (X\hat{w} - y) \in \lambda_n [\partial\Phi(c \circ \cdot)(\hat{w})]_{\hat{J}\setminus J}$$

Note, that $-\frac{X_{\hat{J}\setminus J}^T (X\hat{w}-y)}{\sqrt{n}} = \frac{X_{\hat{J}\setminus J}^T X(\hat{w}-w^*)}{\sqrt{n}} - \frac{X_{\hat{J}\setminus J}^T \epsilon}{\sqrt{n}}$. By CLT, $\frac{X_{\hat{J}\setminus J}^T \epsilon}{\sqrt{n}} \xrightarrow{d} W \sim \mathcal{N}(0, \sigma^2 Q_{\hat{J}\setminus J, \hat{J}\setminus J})$ and by (20) $\hat{w} - w^* \xrightarrow{p} 0$ then $-\frac{X_{\hat{J}\setminus J}^T (X\hat{w}-y)}{\sqrt{n}} = O_p(1)$.

On the other hand, $\frac{\lambda_n c_{\hat{J}\setminus J}}{\sqrt{n}} = \lambda_n n^{\frac{1-\alpha}{2}} n^{\frac{\alpha-1}{2}} c_{\hat{J}\setminus J} \to \infty$, hence $\frac{\lambda_n c_{\hat{J}\setminus J}}{\sqrt{n}} c_{\hat{J}\setminus J}^{-1} v_{\hat{J}\setminus J} \to \infty, \forall v \in \partial\Phi(c \circ \cdot)(\hat{w})$, since $c_{\hat{J}\setminus J}^{-1} v_{\hat{J}\setminus J} = O_p(1)^{-1}$. To see this, let $w'_J = \hat{w}_J$ and 0 elsewhere. Note that by definition of the subdifferential and the stability assumption on J, there must exists $M_J > 0$ s.t

$$\begin{aligned}
\Phi(c \circ w') &\geq \Phi(c \circ \hat{w}) + \langle v_{\hat{J}\setminus J}, -\hat{w}_{\hat{J}\setminus J} \rangle \\
&\geq \Phi(c \circ w') + M_J \|c_{\hat{J}\setminus J} \circ \hat{w}_{\hat{J}\setminus J}\|_\infty - \|c_{\hat{J}\setminus J}^{-1} \circ v_{\hat{J}\setminus J}\|_1 \|c_{\hat{J}\setminus J} \circ \hat{w}_{\hat{J}\setminus J}\|_\infty \\
\|c_{\hat{J}\setminus J}^{-1} \circ v_{\hat{J}\setminus J}\|_1 &\geq M_J
\end{aligned}$$

We deduce then $P(\mathrm{supp}(\hat{w}) \subseteq J) = 1 - P(\hat{J} \setminus J \neq \emptyset) = 1 - P(\text{optimality condition holds}) \to 1$. □

7.4 Discrete stability (Proof of Proposition 2 and relation to weak submodularity)

Proposition 2. *If F is a finite-valued monotone function, F is ρ-submodular iff discrete weak stability is equivalent to strong stability.*

Proof. If F is ρ-submodular and J is weakly stable, then $\forall A \subseteq J, \forall i \in J^c, 0 < \rho[F(J \cup \{i\}) - F(J)] \leq F(J \cup \{i\}) - F(J)$, i.e., J is strongly stable w.r.t. F. If F is such that any weakly stable set is also strongly stable, then if F is not ρ-submodular, then $\forall \rho \in (0, 1]$ there must exists a set $B \subseteq V$, s.t., $\exists A \subseteq B, i \in B^c$, s.t., $\rho[F(B \cup \{i\}) - F(B)] > F(A \cup \{i\}) - F(A) \geq 0$. Hence, $F(B \cup \{i\}) - F(B) > 0$, i.e., B is weakly stable and thus it is also strongly stable and we must have $F(A \cup \{i\}) - F(A) > 0$. Choosing then in particular, $\rho = \min_{B \subseteq V} \min_{A \subseteq B, i \in B^c} \frac{F(A \cup \{i\}) - F(A)}{F(B \cup \{i\}) - F(B)} \in (0, 1]$, leads to a contradiction; $\min_{A \subseteq B, i \in B^c} F(A \cup \{i\}) - F(A) \geq \rho[F(B \cup \{i\}) - F(B)] > F(A \cup \{i\}) - F(A)$. □

We show that ρ-submodularity is a stronger condition than weak submodularity. First, we recall the definition of weak submodular functions.

Definition 7 (Weak Submodularity (c.f., [7, 11])). *A function F is weakly submodular if $\forall S, L, S \cap L = \emptyset, F(L \cup S) - F(L) > 0$,*

$$\gamma_{S,L} = \frac{\sum_{i \in S} F(L \cup \{i\}) - F(L)}{F(L \cup S) - F(L)} > 0$$

Proposition 8. *If F is ρ-submodular then F is weakly submodular. But the converse is not true.*

Proof. If F is ρ-submodular then $\forall S, L, S \cap L = \emptyset, F(L \cup S) - F(L) > 0$, let $S = \{i_1, i_2, \cdots, i_r\}$

$$\begin{aligned}
F(L \cup S) - F(L) &= \sum_{k=1}^{r} F(L \cup \{i_1, \cdots, i_k\}) - F(L \cup \{i_1, \cdots, i_{k-1}\}) \\
&\leq \sum_{k=1}^{r} \frac{1}{\rho} (F(L \cup \{i_k\}) - F(L)) \\
\Rightarrow \gamma_{S,T} &= \rho > 0.
\end{aligned}$$

We show that the converse is not true by giving a counter-example: Consider the function defined on $V = \{1, 2, 3\}$, where $F(\{i\}) = 1, \forall i, F(\{1, 2\}) = 1, F(\{2, 3\}) = 2, F(\{1, 3\}) = 2, F(\{1, 2, 3\}) = 3$. Then note that this function is weakly submodular. We only need to consider sets $|S| \geq 2$, since otherwise $\gamma_{S,T} > 0$ holds trivially. Accordingly, we also only need to consider L which is the empty set or a singleton. In both cases $\gamma_{S,T} > 0$. However, this F is not ρ-submodular, since $F(1, 2) - F(1) = 0 < \rho(F(1, 2, 3) - F(1, 3)) = \rho$ for any $\rho > 0$. □

7.5 Relation between discrete and continuous stability (Proof of Propositions 3 and 4, and Corollary 3)

First, we present a useful simple lemma, which provides an equivalent definition of decomposability for monotone function.

Lemma 7. *Given $w \in \mathbb{R}^d, J \subseteq J, \text{supp}(w) = J$, if Φ is a monotone function, then Φ is decomposable at w w.r.t J iff $\exists M_J > 0, \forall \delta > 0, i \in J^c$, s.t,*

$$\Phi(w + \delta \mathbb{1}_i) \geq \Phi(w) + M_J \delta.$$

Proof. By definition 2, $\exists M_J > 0, \forall \Delta \in \mathbb{R}^d, \text{supp}(\Delta) \subseteq J^c$,

$$\Phi(w + \Delta) \geq \Phi(w) + M_J \|\Delta\|_\infty.$$

in particular this must hold for $\Delta = \delta \mathbb{1}_i$. On the other hand, if the inequality hold for all $\delta \mathbb{1}_i$, then given any Δ s.t. $\text{supp}(\Delta) \subseteq J^c$ let $i_{\max}$ be the index where $\Delta_{i_{\max}} = \|\Delta\|_\infty$ and let $\delta = \|\Delta\|_\infty$, then

$$\Phi(w + \Delta) \geq \Phi(w + \delta_{i_{\max}}) \geq \Phi(w) + M_J \delta = \Phi(w) + M_J \|\Delta\|_\infty.$$

□

Proposition 3. *Given any monotone set function F, all sets $J \subseteq V$ strongly stable w.r.t to F are also strongly stable w.r.t Ω_p and Θ_p.*

Proof. We make use of the variational form (11). Given a set J stable w.r.t to F and $\text{supp}(w) \subseteq J$, let $\kappa^* \in \arg\max_{\kappa \in \mathbb{R}^d_+} \{\sum_{i \in J} \kappa_i^{1/q} |w_i| : \kappa(A) \leq F(A), \forall A \subseteq V\}$, then $\Omega(w) = |w_J|^T (\kappa^*_J)^{1/q}$. Note that $\forall A \subseteq J, F(A \cup i) > F(A)$, by definition 3. Hence, $\forall i \in J^c$, we can define $\kappa' \in \mathbb{R}^d_+$ s.t., $\kappa'_J = \kappa^*_J$, $\kappa'_{(J \cup i)^c} = 0$ and $\kappa'_i = \min_{A \subseteq J} F(A \cup i) - F(A) > 0$. Note that κ' is feasible, since $\forall A \subseteq J, \kappa'(A) = \kappa^*(A) \leq F(A)$ and $\kappa'(A + i) = \kappa^*(A) + \kappa'_i \leq F(A) + F(A \cup i) - F(A) = F(A \cup i)$. For any other set $\kappa'(A) = \kappa'(A \cap (J + i)) \leq F(A \cap (J + i)) \leq F(A)$, by monotonicity. It follows then that $\Omega(w + \delta \mathbb{1}_i) = \max_{\kappa \in \mathbb{R}^d_+} \{\sum_{i \in J \cup i}^d \kappa_i^{1/q} |w_i| : \kappa(A) \leq F(A), \forall A \subseteq V\} \geq |w_J|^T (\kappa^*_J)^{1/q} + \delta (\kappa'_i)^{1/q} \geq \Omega(w) + \delta M$, with $M = (\kappa'_i)^{1/q} > 0$. The proposition then follows by lemma 7.

Similarly, the proof for Θ_p follows in a similar fashion. We make use of the variational form (14). Given a set J stable w.r.t to F and $\text{supp}(w) \subseteq J$, first note that this implicity implies that $F(J) < +\infty$ and hence $\Theta_p(w) < +\infty$. Let $\kappa^* \in \arg\max_{\kappa \in \mathbb{R}^d_+} \sum_{j=1}^d \psi_j(\kappa_j, w_j) + \min_{S \subseteq V} F(S) - \kappa(S)$ and $S^* \in \arg\min_{S \subseteq J} F(S) - \kappa^*(S)$. Note that $\forall S \subseteq J, \forall i \in J^c, F(S \cup i) > F(S)$, by definition 3. Hence, $\forall i \in J^c$, we can define $\kappa' \in \mathbb{R}^d_+$ s.t., $\kappa'_J = \kappa^*_J$, $\kappa'_{(J \cup i)^c} = 0$ and $\kappa'_i = \min_{S \subseteq J} F(S \cup i) - F(S) > 0$. Note that $\forall S \subseteq J, F(S) - \kappa'(S) = F(S) - \kappa^*(S) \geq F(S^*) - \kappa^*(S^*)$ and $F(S+i) - \kappa'(S+i) = F(S+i) - \kappa^*(S) - \kappa'_i \geq F(S+i) - \kappa^*(S) - F(S+i) + F(S) \geq F(S^*) - \kappa^*(S^*)$. Note also that $\psi_i(\kappa'_i, \delta) = (\kappa'_i)^{1/q} \delta$ if $\delta \leq (\kappa'_i)^{1/p}$, and $\psi_i(\kappa'_i, \delta) = \frac{1}{p} \delta^p + \frac{1}{q} \kappa'_i = \delta(\frac{1}{p} \delta^{p-1} + \frac{1}{q} \kappa'_i \delta^{-1}) \geq \delta (\kappa'_i)^{1/q}$ otherwise. It follows then that $\Theta_p(w + \delta \mathbb{1}_i) \geq \sum_{j \in J} \psi_j(\kappa_j, w_j) + (\kappa'_i)^{1/q} \delta + \min_{S \subseteq J \cup i} F(S) - \kappa'(S) \geq \sum_{j \in J} \psi_j(\kappa_j, w_j) + (\kappa'_i)^{1/q} \delta + \min_{S \subseteq J} F(S) - \kappa^*(S) = \Theta_p(w) + \delta M$ with $M = (\kappa'_i)^{1/q} > 0$. The proposition then follows by lemma 7. □

Proposition 4. *If $F = F_-$ and J is strongly stable w.r.t Ω_∞, then J is strongly stable w.r.t F. Similarly, for any monotone F, if J is strongly stable w.r.t Θ_∞, then J is strongly stable w.r.t F.*

Proof. $F(A + i) = \Omega_\infty(\mathbb{1}_A + \mathbb{1}_i) = \Theta_\infty(\mathbb{1}_A + \mathbb{1}_i) > \Omega_\infty(\mathbb{1}_A) = \Theta_\infty(\mathbb{1}_A) = F(A), \forall A \subseteq J$. □

Corollary 3. *If F is monotone submodular and J is weakly stable w.r.t $\Omega_\infty = \Theta_\infty$ then J is weakly stable w.r.t F.*

Proof. If F is a monotone submodular function, then $\Omega_\infty(w) = \Theta_\infty(w) = f_L(|w|)$. If J is not weakly stable w.r.t F, then $\exists i \in J^c$ s.t., $F(J \cup \{i\}) = F(J)$. Thus, given any $w, \text{supp}(w) = J$, choosing $0 < \delta < \min_{i \in J} |w_i|$, result in $f_L(|w| + \delta \mathbb{1}_i) = f_L(|w|)$, which contradicts the weak stability of J w.r.t to $\Omega_\infty = \Theta_\infty$. □