Regularization classes for hypergraphs

September 18, 2023

1 Introduction

The MCP has a wide range of applications (see Bomze et al. [5] and Wu and Hao [24] and references therein) in areas such as social network analysis, telecommunication networks, biochemistry, and scheduling. The MCP is NP hard (Karp [13]). However, owing in part to its wide applicability, a large variety of both heuristic and exact approaches has been investigated (see Bomze et al. [5] for a thorough overview of formulations and algorithms going up to 1999; a more recent survey of algorithms is given in Wu and Hao [24]).

A significant number of the solution methods proposed (for example, Bomze [2], Bomze et al. [6], Gibbons et al. [10], Kuznetsova and Strekalovsky [15], Motzkin and Straus [16], Pelillo [21], and Pelillo and Jagota [22]) are based on solving the following well-known continuous quadratic programming formulation of the MCP due to Motzkin and Straus [16] : max xTAx subject to x , (1)

where is the n-dimensional simplex defined by x Rn : x 0 and 1Tx 1 , and A (aij)i,j9 denotes the adjacency matrix for G defined by aij 1, (i, j)0, (i, j) i, j 9 .

The equivalence between the MCP and (1) is given by the following theorem. 1161 Theorem 1 (Theorem 1 in Motzkin and Straus [16]). The optimal objective value of (1) is 1 1 (G)

, and x(C) is a global maximizer of (1) for any maximum clique C.

Solution approaches to the MCP based on solving (1) include nonlinear programming methods (Gibbons et al. [9] and Pardalos and Phillips [19]) and methods based on discrete time replicator dynamics (Bomze [2], Bomze et al. [5, 6], and Pelillo [21]). Because (1) is NP hard (by reduction to the MCP), the computing time required to obtain a global maximizer can grow exponentially with the size of the graph; hence, finding a global maximizer may be impractical in many settings. However, iterative optimization methods will typically converge to a point satisfying the first-order optimality (Karush–Kuhn–Tucker) conditions. In general, verifying whether a first-order point of a quadratic program is even locally optimal is an NP-hard problem (Murty and Kabadi [17] and Pardalos and Schnitger [20]). However, it was shown in Gibbons et al. [10] that local optimality of a first-order point (in fact, any feasible point) in (1) can be ascertained in polynomial time.

In Pelillo and Jagota [22, proposition 3], a characteristic vector for a clique was shown to satisfy the standard first-order optimality condition for (1) if and only if the associated clique is maximal. In Gibbons et al. [10, theorem 2], the authors gave a characterization of the local optima of (1) and showed a one-one correspondence between strict local maximizers and strictly maximal cliques. These results suggest the possibility of applying iterative optimization methods to (1) to approximately solve the MCP (i.e., to find large maximal cliques). However, one known (Bomze [2], Pelillo [21], and Pelillo and Jagota [22]) drawback of this approach in practice is the presence of “infeasible” or “spurious” local maximizers of (1), which are not characteristic vectors for cliques and from which a clique cannot be recovered through any simple transformation. Such points are an undesirable property of the program, because they can cause continuous-based heuristics to fail by terminating without producing a clique. In Bomze [2], the author addresses this issue by introducing the following regularized formulation (with 1 2):

max xTAx + x2 2 subject to x . (2)

In contrast to (1), the local maximizers of (2) have been shown to be in one-one correspondence with the maximal cliques in G (Bomze [2, theorem 9]), and a replicator dynamics approach to solving (2) was shown to reduce the total number of algorithm failures by 30% compared with a similar approach to solving (1). In Bomze et al. [6], the authors enhanced the algorithm of Bomze [2], adding an annealing heuristic to obtain even stronger results. In addition, it was shown that the correspondence between the local/global optima of (2) and the MCP is maintained for any (0, 1). A similar formulation and approach (Bomze et al. [4]) has also been applied successfully to a weighted version of the MCP. A generalization to hypergraphs was introduced in Rota-Bulo and Pelillo [‘ 23].

In practice, the numerical performance of an iterative optimization method (in terms of speed and/or solution quality) may depend on the particular regularization term used. Authors of [1] relaxed the problem by allowing to take a broad class of regularizers while maintaining full equivalence with the MCP. The goal of this work is to generalize their results to hypergraphs. It is organized as follows : after some definitions and preliminaries in sections 2 and 3, we develop a general regularized formulation of MCP for hypergraphs in section 4 and provide conditions under which the global / local maximizers of the regularized program are in one-one correspondence with the maximum / maximal cliques in G. In section 6, we study some particular subsets of these regularizers and provide some adaptations to the examples given in [1].

2 Definitions and notations

A *k-graph* is a pair $G := (V, E)$, where $V := [1, \dots, n]$ is a finite set of *vertices* and $E \subseteq \binom{V}{k}$ is a set of k-subsets of V, each of which is called a *hyperedge*. 2-graphs are typically called *graphs*. We denote by $\overline{G} := (V, \overline{E} := \binom{V}{k} \setminus E)$ its complementary.

The lagrangian of graph G L_G is defined such that $L_G(x) := \sum_{e \in E} \prod_{i \in e} x_i$.

We denote $\Delta := \{x \in \mathbb{R}^n | x \geq 0 \text{ and } \mathbf{1}^T x = 1\}$ the n-dimensional simplex. We also define the following set :

$$\Delta^0 := \underset{C \text{ clique}}{\cup} \Delta^{(C)} = \{x \in \Delta | \text{supp}(x) \text{ is a clique}\} \tag{1}$$

We define the set of permutations of x as

$$\mathcal{P}(x) := \left\{\overline{x} | \exists \sigma \in \mathcal{S}_n, \, \overline{x}_{\sigma(i)} = x_i \; \forall i \in [1, \dots, n]\right\} \tag{2}$$

We denote $\mathbf{1}_C$ the characteristic function with logical expression C i.e. $\mathbf{1}_C = 0$ when C is false and $\mathbf{1}_C = 1$ when C is true, and $\mathbf{e}_i$ the i-th column of the identity matrix i.e. a vector with a 1 at index i and 0 everywhere else.

A characteristic vector of a subset A is defined by $x^{(A)} := (\frac{\mathbf{1}_{i \in A}}{|A|})_i \in \Delta$.

For every $x \in \Delta$, $i \in [1, \dots, n]$, the multiplier function $\lambda_i : \Delta \to \mathbb{R}$ for a function f is defined as

$$\lambda_i(x) := \nabla f(x)^T (\mathbf{e}_i - x) \tag{3}$$

or in vector form

$$\lambda(x) := \nabla f(x) - x^T \nabla f(x) \mathbf{1} \tag{4}$$

3 Previous work

The goal of this work is to generalize to hypergraphs the results found in [1], which are briefly summarized hereafter :

Consider the following problem :

$$\max_{\text{s.t. } x \in \Delta} x^T A x + \phi(x) \tag{5}$$

where ϕ satisfies the following conditions :

- $\nabla^2\phi(x) \geq 0$ ie ϕ is convex
- $\|\nabla^2\phi(x)\|_2 < 2$
- $\phi(x)$ is invariant by permutations of x

The global and local maximizers of the above problem are characteristic vectors of respectively maximum and maximal cliques.

Examples of such regularizers are :

- $\phi(x) = \frac{1}{2}\|x\|_2^2$
- $\alpha_1\|x + \epsilon 1\|_p^p$ with $\epsilon \geq 0$, $p \geq 2$ and $0 \leq \alpha_1 \leq \frac{2}{p(p-1)(1+\epsilon)^{p-2}}$
- $\phi(x) = \alpha_2 \sum_{i=1}^n (e^{-\beta x_i} - 1)$, with $\beta > 0$ and $0 \leq \alpha_2 \leq \frac{2}{\beta^2}$ (approximation of $-\alpha_2\|x\|_0$)

4 Formulation considered

Let $G := (V, E)$ be a k-graph with vertices V and edges E and consider the following problem :

$$\min_{\text{s.t. } (x,y)\in\Delta} L_{\overline{G}}(x) + \Phi(x) \tag{6}$$

where Φ satisfies the three following assumptions :

(C1) $\Phi \in \mathcal{C}^2(\mathbb{R}^n)$ and for all face S of Δ, $\nabla^2\Phi_S(x) > 0$ for all $x \in \Delta^{(S)}$ i.e. the restriction of Φ to any face of Δ is strictly convex

(C2) $\Phi(\overline{x}) = \Phi(x)$ for all $\overline{x} \in \mathcal{P}(x)$ i.e. Φ is symmetric / permutation invariant

(C3) $(\mathbf{e}_i - \mathbf{e}_j)^T\nabla^2\Phi(x)(\mathbf{e}_i - \mathbf{e}_j) < x_i^{k-2} + x_j^{k-2}$ for every $x \in \Delta$, $i, j \in \text{supp}(x)$ with $i \neq j$

In the following, we will denote $h(x) := L_{\overline{G}}(x) + \Phi(x)$. We claim that minimizers of this problem are strict and attained at $x = x^{(C)}$ with C a maximal clique in G.

Remark 4.1. *For $k = 2$, we recover a condition **(C3)** similar to condition **(C3)** from [1].*

5 Proof

Consider the following problem, where $C \subseteq V$ is any clique :

$$\max_{\text{s.t } x\in\Delta^{(C)}} h(x) \tag{7}$$

Proposition 5.1. *The unique local (hence, global) minimizer of (7) is $x^{(C)}$.*

Proof. For $x \in \Delta^{(C)}$, $h(x) = L_{\overline{G}}(x) + \Phi(x) = \Phi_C(x_C)$ because $\text{supp}(x) = C$ is a clique. As $\Delta^{(C)}$ is convex and Φ_C is strictly convex, there exists a unique minimizer of h on $\Delta^{(C)}$, which is necessarily $x^{(C)}$. Indeed, denote x^* this minimizer. As Φ is permutation invariant, we must have $\mathcal{P}(x^*) \cap \Delta^{(C)} = \{x^*\}$ i.e. all the indices in C are equal, and because $\sum_i x_i = 1$, necessarily $x^* = x^{(C)}$. $\square$

The previous proposition shows that the unique maximizer in a face whose support is a clique is the characteristic vector of that clique. The next step is to prove that when we join these faces whose supports are cliques, the only maximizers remaining are the characteristic vectors of maximal cliques. Thus consider the following problem :

$$\min_{\text{s.t. } (x,y)\in\Delta^0} h(x) \tag{8}$$

Proposition 5.2. *A point $x \in \Delta^0$ is a local minimizer of (8) if and only if $x = x^{(C)}$ for some maximal clique C. Moreover, every local minimizer of (8) is strict.*

Proof. $\Rightarrow$: First, observe that, for any local minimizer x of (8), by definition of Δ^0, there exists some maximal clique C such that $x \in \Delta^{(C)}$ which implies that $x = x^{(C)}$ by the previous proposition.

$\Leftarrow$: Let C be a maximal clique, and suppose by way of contradiction that $x^{(C)}$ is not a strict minimizer in (8). Then, for every $k \in \mathbb{N}^*$, there exists some $x^k \in \Delta^0$ with $0 < \|x^k - x^{(C)}\|_2 < 1/k$ such that $h(x^k) \leq h(x^{(C)})$. Because there are only finitely many sets in the union of (1), there must exist some clique C' and some subsequence $(x^{k_l})_{l=1}^{+\infty} \subseteq (x^k)_{k=1}^{+\infty}$ such that $x^{k_l} \in \Delta(C')$ for each $l \geq 1$, with $x^{k_l} \to x^{(C)}$. Hence, $x^{(C)} \in \overline{\Delta(C')} = \Delta(C')$, which implies $C = \text{supp}(x^{(C)}) \subseteq C'$. Because C is maximal, we must have $C' = C$ and thus $x^{k_l} \in \Delta(C') = \Delta^{(C)}$ for each $l \geq 1$. Thus $x^{(C)}$ is not a strict local maximizer of (7), contradicting 5.1. $\square$

The next lemma is essential in order to prove the correspondence between global maxima of the continuous problem and the maximum cliques of the graph.

Lemma 5.3. *If C^1 and C^2 are cliques, then*

$$|C^1| < |C^2| \iff h(x^{(C^1)}) > h(x^{(C^2)}) \tag{9}$$

Proof. Let C^1, C^2 be cliques.

$\Rightarrow$: Assume that $|C^1| < |C^2|$. Let C be any clique such that $C \subset C^2$ and $|C| = |C^1|$. Then $x^{(C^1)} \in \mathcal{P}(x^{(C)})$. Therefore, $h(x^{(C^1)}) = h(x^C)$.

Moreover by 5.1, $h(x^{(C)}) > h(x^{(C^2)})$, because $x^{(C^2)} \in \Delta(C^2)$. Hence, $h(x^{(C^1)}) > h(x^{(C^2)})$.

$\Leftarrow$: Assume that $h(x^{(C^1)}) > h(x^{(C^2)})$. By contraposition of the previous part, we must have $|C^1| \leq |C^2|$. Moreover, if $|C^1| = |C^2|$, then $x^{(C^1)} \in \mathcal{P}(x^{(C^2)})$ and thus $h(x^{(C^1)}) = h(x^{(C^2)})$, a contradiction. $\square$

The proof of the correspondence is now straightforward.

Proposition 5.4. *A point $x \in \Delta^0$ is a global minimizer of (8) if and only if $x = x^{(C)}$ for some maximum clique C.*

Proof. Let $x \in \Delta^0$. Then, x is a global minimizer if and only if x is a local minimizer and $h(x) \leq h(\tilde{x})$ for every local minimizer $\tilde{x} \neq x$, which by 5.2, holds if and only if $x = x^{(C)}$ for some maximal clique C and $h(x^{(C)}) \leq h(x^{\tilde{C}})$ for every maximal clique $\tilde{C} \neq C$. The corollary then follows from the previous proposition. $\square$

So far, we have proved that the local and global maxima are characteristic vectors of respectively maximal and maximum cliques only on the union of the face whose supports are cliques. The last step is to show that there is no maximizer lying outside these faces, and this is where condition **(C3)** becomes necessary.

Proposition 5.5. *Every local minimizers of (6) are in Δ^0.*

Proof. Let x be a minimizer of (6). If $x \in \Delta^0$ we are done. Therefore assume instead that $x \notin \Delta^0$. Then $\text{supp}(x)$ is not a clique, and there exists an edge $\tilde{e} \in \overline{E}$ such that for all $k \in \tilde{e}$, $k \in \text{supp}(x)$. Let $i, j \in \tilde{e}$ such that $x_i \leq x_j \leq \min_{k \in \tilde{e} \setminus \{i,j\}} x_k$ and define $x(t) := x + t(\mathbf{e}_i - \mathbf{e}_j)$ for $-x_i \leq t \leq x_j$.

Because $d = t(\mathbf{e}_i - \mathbf{e}_j)$ is feasible for $-x_i \leq t \leq x_j$, we have that $\nabla h(x)^T d = 0$, and taking a Taylor's expansion around x we get

$$\begin{aligned} h(x(t)) - h(x) &= \frac{1}{2} d^T \nabla^2 h(x) d + \mathcal{O}(t^3) && (10) \\ &= \frac{t^2}{2} \left[-2 \sum_{e \in \overline{E}} 1_{i,j \in e} \prod_{l \in e \setminus \{i,j\}} x_l + (\mathbf{e}_i - \mathbf{e}_j)^T \nabla^2 \Phi(x) (\mathbf{e}_i - \mathbf{e}_j) \right] + \mathcal{O}(t^3) && (11) \end{aligned}$$

First,

$$-2 \sum_{e \in \overline{E}} 1_{i,j \in e} \prod_{l \in e \setminus \{i,j\}} x_l \leq -x_i^{k-2} - x_j^{k-2} \tag{12}$$

because at least edge $\tilde{e}$ contributes to one non-null term in the sum and all of the x_l are greater than x_i and x_j, and then by assumption **(C3)**

$$(\mathbf{e}_i - \mathbf{e}_j)^T \nabla^2 \Phi(x)(\mathbf{e}_i - \mathbf{e}_j) < x_i^{k-2} + x_j^{k-2} \tag{13}$$

Thus $h(x(t)) - h(x) < 0$ when t is sufficiently small i.e. there is no neighborhood of x in which $h(x)$ is a minimum thus x is not a minimizer, a contradiction. $\square$

We have that local and global maximizers are characteristic vectors of respectively maximal and maximum cliques, we now have to prove that all of these characteristic vectors are themselves maximizers.

Proposition 5.6. *Let C be a maximal clique, then $x^{(C)}$ is a local minimizer of* (6).

Proof. $x^{(C)}$ is a local minimizer if and only if strict complementarity stands, i.e. $\lambda_i(x^{(C)}) = 0$ for $i \in \operatorname{supp}(x^{(C)})$ and $\lambda_i(x^{(C)}) > 0$ for $i \notin \operatorname{supp}(x^{(C)})$.

Case 1 : $C = V$ i.e. the clique is the whole graph. As there is no $i \notin C$, it is sufficient to check that $\lambda_i(x^{(C)}) = 0$ for all i. Let $i \in C = \operatorname{supp}(x^{(C)})$, we have

$$\lambda_i(x^{(C)}) = \left(\nabla h(x^{(C)})\right)_i - \nabla h(x^{(C)})^T x^{(C)} = \left(\nabla h(x^{(C)})\right)_i - \sum_{l \in C} \left(\nabla h(x^{(C)})\right)_l x_l^{(C)} \tag{14}$$

$$= \left(\nabla h(x^{(C)})\right)_i - \left(\nabla h(x^{(C)})\right)_i \sum_{l \in C} x_l^{(C)} \tag{15}$$

$$= 0 \tag{16}$$

where we used the symmetry of Φ and $x_i^{(C)} = x_l^{(C)}$ for all l in (15) and $\sum_l x_l = 1$ for all $x \in \Delta$ in (16), thus we do have strict complementarity in this case.

Case 2 : $C \neq V$ i.e. there is at least one missing node. In this case, strict complementarity is equivalent to

$$\nabla h(x^{(C)})^T(\mathbf{e}_i - \mathbf{e}_j) > 0 \text{ for } i \notin C \text{ and } j \in C \tag{17}$$

Indeed, $\lambda_i(x^{(C)}) = \nabla h(x^{(C)})^T(\mathbf{e}_i - x^{(C)})$ so $\nabla h(x^{(C)})^T(\mathbf{e}_i - \mathbf{e}_j) = \lambda_i(x^{(C)}) - \lambda_j(x^{(C)})$ and for $j \in C = \operatorname{supp}(x^{(C)})$, we already have $\lambda_j(x^{(C)}) = 0$ by the same proof as for **Case 1**.

Now let $i \notin C$ and $j \in C$ and define $\tilde{x} := x^{(C)} + (e_i - e_j)/(2|C|)$. Then

$$\nabla L_{\overline{G}}(\tilde{x})^T(e_i - e_j) = \sum_{e \in \overline{E}} \mathbf{1}_{i \in e} \prod_{l \in e \setminus \{i\}} \tilde{x}_l - \sum_{e \in \overline{E}} \mathbf{1}_{j \in e} \prod_{l \in e \setminus \{j\}} \tilde{x}_l > 0 \tag{18}$$

because $i \notin C$ and there is at least one edge in $\overline{E}$ between $k-1$ vertices of C and i (or else the clique would not be maximal), and $j \in C$ so the second sum is null. Then by condition **(C2)** we get

$$\nabla \Phi(\tilde{x})^T(e_i - e_j) = \frac{\partial h}{\partial x_i}(\tilde{x}) - \frac{\partial h}{\partial x_j}(\tilde{x}) = \frac{\partial h}{\partial x_i}(\tilde{x}) - \frac{\partial h}{\partial x_i}(\tilde{x}) = 0 \tag{19}$$

because $\tilde{x}_i = \tilde{x}_j$.

Now define $\overline{h}(t) := h(x(t))$ where $x(t) := x^{(C)} + t(\mathbf{e}_i - \mathbf{e}_j)$ for $0 \le t \le x_j = 1/|C|$, so that $\overline{h}'(1/2|C|) = \left(\nabla L_{\overline{G}}(\tilde{x}) + \Phi(\tilde{x})\right)^T (\mathbf{e}_i - \mathbf{e}_j) > 0$. For $0 \le t \le x_j = 1/|C|$, we have

$$\overline{h}''(t) = (\mathbf{e}_i - \mathbf{e}_j)^T \nabla^2 h(x(t))(\mathbf{e}_i - \mathbf{e}_j) \tag{20}$$

$$= -2 \sum_{e \in \overline{E}} 1_{i,j \in e} \prod_{l \in e \setminus \{i,j\}} x(t)_l + (\mathbf{e}_i - \mathbf{e}_j)^T \nabla^2 \Phi(x(t))(\mathbf{e}_i - \mathbf{e}_j) \tag{21}$$

$$< 0 \tag{22}$$

by the same reasoning as in the proof of 5.5. Thus $\overline{h}'$ is strictly decreasing, so $\overline{h}'(0) > \overline{h}'(1/2|C|) > 0$. This concludes the proof since $\nabla h(x^{(C)})^T(\mathbf{e}_i - \mathbf{e}_j) = \overline{h}'(0)$. □

The correspondence "local / global maximizer $\longleftrightarrow$ characteristic vector of maximal / maximum cliques" is now established.

6 Particular cases

This section is dedicated to the simplification of conditions **(C1) - (C3)** in various particular cases of the regularizer.

6.1 Sum of real functions

Assume $k > 2$ and that $\Phi(x) = \sum_{i=1}^n q_i(x_i)$ for some $q_i \in \mathcal{C}^2(\mathbb{R}, \mathbb{R})$ i.e. that the regularizer is a sum of twice derivable real functions depending on only one dimension of x.

Lemma 6.1. *All the q_i are equal up to a constant, i.e. there exists $q : \mathbb{R} \to \mathbb{R}$ such that for all $1 \le i \le n$, $q_i' = q'$.*

Proof. Follows from condition **(C2)**. □

This lemma shows that necessarily, Φ can be rewritten as $\Phi(x) = \sum_{i=1}^n q(x_i) + C$ where C is a constant.

Lemma 6.2. *Necessarily, $q''(0) = 0$.*

Proof. By condition **(C1)**, $q \in \mathcal{C}^2(\mathbb{R})$ and $q''(y) > 0$ when $y > 0$. Taking the limit $x \to 0$, we get $q''(0) \ge 0$. Now for $0 < t \le \frac{1}{2}$ define $x(t) := (t, t, 1-2t, 0, \dots, 0) \in \Delta$. By condition **(C3)**, we have $2q''(t) \le 2t^{k-2}$. Taking the limit $t \to 0$ gives us $q''(0) \le 0$ thus $q''(0) = 0$. □

The next proposition simplifies condition **(C3)**, allowing to study the regularizer on only one dimension instead of two at a time.

Proposition 6.3. *In this case, conditions **(C1) - (C3)** can be rewritten as*

- ***(C'1)** $q \in \mathcal{C}^2(\mathbb{R})$ and $q''(y) > 0$ when $y > 0$*
- ***(C'2)** $q''(y) < y^{k-2}$ for $0 < y \le \frac{1}{2}$ and $q''(y) \le y^{k-2}$ for $\frac{1}{2} < y \le 1$*

Proof. $\Leftarrow$: if q satisfies **(C'1)** and **(C'2)**, then $\Phi(x) = \sum_{i=1}^n q(x_i)$ obviously satisfies **(C1) - (C3)**.

$\Rightarrow$: Assume Φ satisfies **(C1) - (C3)**. Then for all $x \in \Delta$, $\frac{\partial^2 \Phi}{\partial x_i^2}(x) = q''(x_i)$ so q satisfies **(C'1)**.

Let $y \in (0,1)$ and for $0 \le t < 1-y$ define $x(t) := (y, 1-y-t, t, 0, \dots, 0) \in \Delta$. By **(C3)**, we have that for all t, $q''(y) + q''(1-y-t) < y^{k-2} + (1-y-t)^{k-2}$. Taking the limit $t \to 1-y$, we get $q''(y) \le y^{k-2}$. This inequality can be extended to $y = 1$ by continuity of q''. Now assume by contradiction there exists $y \in (0, \frac{1}{2}]$ such that $q''(y) = y^{k-2}$. Then applying **(C3)** with $x = (y, y, 1-2y, 0, \dots, 0) \in \Delta$, we get $2y^{k-2} = 2q''(y) < 2y^{k-2}$, a contradiction. □

6.2 Generalization of the regularizers found in [1]

This section is dedicated to the adaptation of the three original regularizers given in [1], which were :

- $\phi(x) = \frac{1}{2}\|x\|_2^2$
- $\alpha_1\|x+\epsilon 1\|_p^p$ with $\epsilon \geq 0$, $p \geq 2$ and $0 \leq \alpha_1 \leq \frac{2}{p(p-1)(1+\epsilon)^{p-2}}$
- $\phi(x) = \alpha_2 \sum_{i=1}^n (e^{-\beta x_i} - 1)$, with $\beta > 0$ and $0 \leq \alpha_2 \leq \frac{2}{\beta^2}$ (approximation of $-\alpha_2\|x\|_0$)

We claim that the three following regularizers satisfy conditions **(C1)** - **(C3)** : NOT FINISHED

- $\Phi_B(x) := \alpha\|x\|_k^k$ with $0 < \alpha < \frac{1}{k(k-1)}$
- $\Phi_1(x) := \alpha_1\|x+\epsilon\mathbf{1}\|_p^p - \frac{n}{2}\alpha_1 p(p-1)\epsilon^{p-2}x^2$ with $\epsilon > 0$, $p \geq k$ and $0 < \alpha_1 \leq \frac{1}{p(p-1)(1+\epsilon)^{p-2}}$
- $\Phi_2(x) := \alpha_2 \sum_{i=1}^n \left(e^{-\beta x_i} - 1\right)$ with $\beta > 0$, and $0 < \alpha_2 < \frac{1}{\beta^2}$

Proof. These three regularizers are sum of twice derivable functions depending on only one dimension of x, the results of the previous part can be applied.

- Φ_B : Obvious.
- Φ_1 : For all $x \in \Delta$, $\Phi_1(x) = \sum_{i=1}^n \left(\alpha_1(x_i+\epsilon)^p - \frac{1}{2}\alpha_1 p(p-1)\epsilon^{p-2}x^2\right)$. For $y \in [0,1]$, denote $q(y) = \alpha_1(y+\epsilon)^p - \frac{1}{2}\alpha_1 p(p-1)\epsilon^{p-2}x^2$. $q''(y) = \alpha_1 p(p-1)(y+\epsilon)^{p-2} - \alpha_1 p(p-1)\epsilon^{p-2} > 0$ when $y > 0$ and for $y > 0$, $q''(y) < \alpha_1 p(p-1)(y^{p-2}+\epsilon^{p-2}) - \alpha_1 p(p-1)\epsilon^{p-2} = \alpha_1 p(p-1)y^{p-2} \leq y^{k-2}$.
- Φ_2 : For $y \in [0,1]$, denote $q(y) = \alpha_2(e^{-\beta y} - 1)$. $q''(y) = \alpha_2\beta^2 e^{-\beta y}$

□

Remark 6.4. *The regularizer Φ_2 is the best approximation under the norm $\|\cdot\|_2$ of $y \to \alpha_2(e^{-\beta y} - 1)$ with the form $y \to ay^l + by + c$ under the constraints $l \geq k$ and $0 < a \leq \frac{1}{l(l-1)}$. Indeed, the problem can be reformulated as*

$$\begin{aligned} &\min_{a,b,c} \int_0^1 \left(ay^l + by + c - \alpha_2(e^{-\beta y} - 1)\right)^2 dy \\ &s.t.\ l \geq k \text{ and } 0 < a \leq \frac{1}{l(l-1)} \end{aligned} \tag{23}$$

Then the KKT stationarity conditions give :

$$0 = \int_0^1 2y^l \left(ay^l + by + c - \alpha_2(e^{-\beta y} - 1)\right) dy - \mu_1 + \mu_2 \tag{24}$$

$$= \frac{2a}{2l+1} + \frac{2b}{l+2} + \frac{2c}{l+1} - \frac{2\alpha_2}{l+1} - 2\alpha_2 \int_0^1 y^l e^{-\beta y} dy + \frac{2\alpha_2}{l+1} - \mu_1 + \mu_2 \tag{25}$$

$$= \frac{2a}{2l+1} + \frac{2b}{l+2} + \frac{2c}{l+1} - \frac{2\alpha_2}{l+1} - \frac{2\alpha_2}{\beta^{l-1}}\left[\Gamma(l+1,0) - \Gamma(l+1,\beta)\right] + \frac{2\alpha_2}{l+1} - \mu_1 + \mu_2 \tag{26}$$

$$0 = \int_0^1 2y \left(ay^l + by + c - \alpha_2(e^{-\beta y} - 1)\right) dy = \frac{2a}{l+1} + b + 2c + \frac{2\alpha_2}{\beta}e^{-\beta} + \frac{2\alpha_2}{\beta^2}e^{-\beta} - \frac{2\alpha_2}{\beta^2} - 2\alpha_2 \tag{27}$$

$$0 = \int_0^1 2\left(ay^l + by + c - \alpha_2(e^{-\beta y} - 1)\right) dy = \frac{2a}{l} + 2b + 2\frac{\alpha_2}{\beta} - 2\frac{\alpha_2}{\beta}e^{-\beta} \tag{28}$$

$$0 = \int_0^1 2a\ln(y)y^l \left(ay^l + by + c - \alpha_2(e^{-\beta y} - 1)\right) dy - \mu_l \tag{29}$$

where we used an integration by substitution in the first equality and an integration by parts in the middle equality. The other conditions also give :

$$0 < a \leq \frac{1}{l(l-1)} \tag{30}$$

$$k \leq l \tag{31}$$

$$\mu_1 = 0 \tag{32}$$

$$\mu_2, \mu_l \geq 0 \tag{33}$$

$$\mu_2 \left(a - \frac{1}{l(l-1)} \right) = 0 \tag{34}$$

$$\mu_l (k - l) = 0 \tag{35}$$

References

[1] James T. Hungerford and Francesco Rinaldi. *A General Regularized Continuous Formulation for the Maximum Clique Problem*. 2017. arXiv: 1709.02486 [math.OC].