MIQCQP reformulation of the ReLU neural networks Lipschitz constant estimation problem

Mohammed Sbihi, Sophie Jan, Nicolas Couellan

▶ **To cite this version:**

Mohammed Sbihi, Sophie Jan, Nicolas Couellan. MIQCQP reformulation of the ReLU neural networks Lipschitz constant estimation problem. 2024. hal-04431914

HAL Id: hal-04431914

https://enac.hal.science/hal-04431914

Preprint submitted on 1 Feb 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Public Domain

MIQCQP reformulation of the ReLU neural networks Lipschitz constant estimation problem

Mohammed Sbihi*[1], Sophie Jan†[2], and Nicolas Couellan‡[1,2]

[1]Fédération ENAC ISAE-SUPAERO ONERA, Université de Toulouse, France
[2]Institut de Mathématiques de Toulouse UMR 5219, Université de Toulouse; CNRS, UPS, F-31062 Toulouse Cedex 9, France.

Abstract

It is well established that to ensure or certify the robustness of a neural network, its Lipschitz constant plays a prominent role. However, its calculation is NP-hard. In this note, by taking into account activation regions at each layer as new constraints, we propose new quadratically constrained MIP formulations for the neural network Lipschitz estimation problem. The solutions of these problems give lower bounds and upper bounds of the Lipschitz constant and we detail conditions when they coincide with the exact Lipschitz constant.

1 Introduction

Several studies have demonstrated the prominent role of the Lipschitz constant in the robustness of neural networks. It has been shown for example that it is related to generalization bounds of neural network classifiers [11, 1]. The Lipschitz constant expresses also the maximum variation of the neural network outputs and can therefore be used to derive robustness certificates when inputs are subject to random or adversarial perturbations [10]. Furthermore, it has been used as a regularization term of the neural network training loss function to compute optimal neural network weights that achieve better robustness [5]. Alternatively, constraints on the Lipschitz constant have been added in the training loss minimization problem to develop 1-Lipschitz neural networks [2].

The exact calculation of the Lipschitz constant of a neural network is a difficult problem. Even in the simple case of one hidden layer neural network, it can be shown that the problem is NP-hard [10]. Therefore, an estimation of the constant in the form of upper bounds (sometimes lower bounds) is usually sought. However, in the common case of ReLU networks, computing tight estimates of the constant is also an NP-hard problem (see Theorem 4 in [9]). Several approaches have been proposed in the literature. They vary by the scope of the input being considered (global or local Lipschitz regularity), the order p of L_p-Lipschitz regularity, or the underlying estimation method.

Among these various approaches, Lipschitz certificates via semidefinite programming (SDP) are proposed in [4]. By exploiting slope restriction properties of common activation functions, incremental quadratic constraints are formulated and the global L_2-Lipschitz constant problem is then expressed as a SDP. While it provides a nice convex formulation of the constant estimation, its application to real life neural networks architectures is limited by the computational complexity of available methods for solving SDP. Alternatively, in the case of ReLU networks, authors have considered mixed integer programming (MIP) approaches to derive exact or upper bounds of the Lipschitz constant. Indeed,

*mohammed.sbihi@recherche.enac.fr
†sophie.jan@math.univ-toulouse.fr
‡nicolas.couellan@recherche.enac.fr

in [9], by showing that a ReLU network is a composition of MIP-encodable components and therefore itself MIP-encodable, the authors formulate the exact calculation of the local Lipschitz constant of ReLU networks as a MIP. In the worst case, MIP problems have exponential time complexity, however, in practice they are often solved in reasonable time.

In this note, we propose new quadratically constrained MIP formulations for the neural network Lipschitz estimation problem. First, by taking into account activation regions at each layer as new constraints, we derive three new MIP formulations whose solutions give a lower bound $\underline{L}$ of the Lipschitz constant, a sequence $\{\underline{L}_\epsilon\}_{\epsilon>0}$ of lower bounds converging to $\underline{L}$ and an upper bound $\bar{L}$. We further show that $\underline{L}$ and $\bar{L}$ coincide and are equal to the true Lipschitz constant if the neural network is in general position as defined in [9]. We also show that, except on a set of network parameters of Lebesgue measure 0, $\underline{L}$ coincide with $\bar{L}$. Next, by reformulating the activation constraints as quadratic constraints, we propose equivalent Mixed Integer Quadratically Constrained Quadratic Program (MIQCQP) formulations. These new constrained problems have the benefit of reducing the search space in the branching phase involved in the MIP solving process. Furthermore, the specific quadratic structure of the objective and the constraints can also be exploited in the bounding phase of MIP solvers using quadratic convex relaxations and linearizations strategies as explained in [3]. However, the study of the numerical solutions of these problems goes beyond the scope of this note that was only intended to explain the derivation of the MIQCQP formulation of the neural network Lipschitz estimation problem.

The note is organized in three sections. Section 2 introduces the general problem of calculating the Lipschitz constant $L(f, \mathcal{X})$ of a neural network f over an input set $\mathcal{X}$. Section 3 details the derivation of lower and upper bounds for $L(f, \mathcal{X})$. Finally, Section 4 provides Mixed Integer Quadratically Constrained Quadratic Program reformulations for the problem of estimating $L(f, \mathcal{X})$ using the bounds obtained in Section 3.

2 Problem statement

We consider ReLU Multi-Layer-Perceptron function $f : \mathbb{R}^{n_0} \to \mathbb{R}^{n_L}$, that is a composition of affine operators and element-wise ReLU nonlinearities. More precisely, it may be encoded by:

$$f(x) = T_L \circ \rho_{L-1} \circ T_{L-1} \circ \cdots \circ \rho_1 \circ T_1(x)$$

where $T_k : \mathbb{R}^{n_{k-1}} \ni x \mapsto M_k x + b_k \in \mathbb{R}^{n_k}$ is an affine function and $\rho_k : \mathbb{R}^{n_k} \to \mathbb{R}^{n_k}$ is the ReLU operator applied element-wise. We denote by $\theta_k = T_k \circ \rho_{k-1} \circ T_{k-1} \circ \cdots \circ \rho_1 \circ T_1$ the pre-activation output of the k-th layer and by θ_k^i the i-th component of θ_k (corresponding to the pre-activation of the i-th neuron of layer k). In the following, given an element $v \in \mathbb{R}^n$, we denote its $i-th$ component by v^i and we denote Hadamard product between two vectors v_1 and v_2 by $v_1 \odot v_2$.

We are interested in computing the quantity

$$\sup_{x,y \in \mathcal{X}} \frac{\|f(y) - f(x)\|}{\|y - x\|},$$

where $\mathcal{X}$ is an open subset of $\mathbb{R}^n$ and $\|.\|$ is a norm. When this quantity is finite, we denote it by $L(f, \mathcal{X})$ and we say that f is locally Lipschitz over $\mathcal{X}$. If $\mathcal{X} = \mathbb{R}^{n_0}$, then we denote the above quantity $L(f)$ and we simply say that f is (globally) Lipschitz.

3 Deriving lower and upper bounds of the Lipschitz constant of ReLU networks

We now derive an upper and lower bounds for $L(f, \mathcal{X})$ by observing [9, Theorem 1] that

$$L(f, \mathcal{X}) = \text{esssup}_{x \in \mathcal{X}} \sup_{G \in \mathcal{J}f(x)} \|G\| \tag{1}$$

where $\mathcal{J}f(x)$ is the (Clarke) generalized jacobian of f at x. Using recursively the Clarke jacobian Chain Rule (see [8, Theorem 4]), we obtain the following bound

$$L(f,\mathcal{X}) \leq \sup_{x\in\mathcal{X}} \left\{ \|M_L \mathrm{diag}(g_{L-1})M_{L-1}\cdots \mathrm{diag}(g_1)M_1\| \quad | \quad g_k \in [0,1]^{n_k},\ g_k^i \in \partial\mathrm{ReLU}(\theta_k^i(x)) \right\}. \quad (2)$$

Here $\partial\mathrm{ReLU}(x)$ is the subdifferential of the ReLU function:

$$\partial\mathrm{ReLU}(x) = \begin{cases} \{0\} & \text{if } x<0, \\ [0,1] & \text{if } x=0, \\ \{1\} & \text{if } x>0. \end{cases}$$

An activation pattern for the ReLU network f is an assignment to each hidden neuron of a sign 1 or 0:

$$(\sigma_1,\sigma_2,\ldots,\sigma_{L-1}) \in \{0,1\}^{n_1}\times\{0,1\}^{n_2}\times\cdots\times\{0,1\}^{n_{L-1}}.$$

The activation region in $\mathcal{X}$ corresponding to $(\sigma_1,\sigma_2,\ldots,\sigma_{L-1})$ is

$$\mathcal{R}\left(\mathcal{X};(\sigma_1,\sigma_2,\ldots,\sigma_{L-1})\right) = \left\{ x\in\mathcal{X} \quad | \quad \left(\sigma_k^i - \frac{1}{2}\right)\theta_k^i(x) > 0,\ \forall i\in\{1,\ldots,n_k\}, \forall k\in\{1,\cdots,L-1\} \right\}.$$

When specifying an activation pattern, the signal assigned to a neuron i from the $k-th$ layer determines whether it is on or off for inputs in the activation region since the pre-activation of neuron is positive (resp. negative) when $\sigma_k^i = 1$ (resp. $\sigma_k^i = 0$). Let us define $H_{k,i} = \{x\in\mathbb{R}^{n_0} \mid \theta_k^i(x) = 0\}$ (for $i\in\{1,\ldots,n_k\}, k\in\{1,\cdots,L-1\}$) which can be thought of as "bent hyperplanes". The non-empty activation regions is not else but the connected components of $\mathcal{X}\setminus\cup_{k,i}H_{k,i}$ (Lemma 2 in [6]). The Jacobian exists and is the same for all the points belonging to the same activation pattern. The Jacobian corresponding to a pattern $(\sigma_1,\cdots,\sigma_{L-1})$ is equal to $M_L\mathrm{diag}(\sigma_{L-1})M_{L-1}\cdots\mathrm{diag}(\sigma_1)M_1$. We obtain the following lower bound for $L(f,\mathcal{X})$:

$$L(f,\mathcal{X}) \geq \sup_{(\sigma_1,\sigma_2,\ldots,\sigma_{L-1})\in\{0,1\}^{n_1+n_2+\cdots+n_{L-1}} | \mathcal{R}(\mathcal{X};\sigma)\neq\emptyset} \|M_L\mathrm{diag}(\sigma_{L-1})M_{L-1}\cdots\mathrm{diag}(\sigma_1)M_1\|. \quad (3)$$

Denoting $N(g) = \|M_L\mathrm{diag}(g_{L-1})M_{L-1}\cdots\mathrm{diag}(g_1)M_1\|$, we are thus interested in solving the following two optimization problems to obtain upper and lower bounds of the Lipshitz constant of the ReLU neural network:

$$(\mathcal{U}^*)\begin{cases} \max & N(g) \\ \text{s.t.} & g_k^i \in \partial\mathrm{ReLU}(\theta_k^i(x)),\ \forall i\in\{1,\ldots,n_k\}, k\in\{1,\cdots,L-1\} \\ & g_k\in[0,1]^{n_k}, \forall k\in\{1,\cdots,L-1\} \\ & x\in\mathcal{X} \end{cases}$$

and

$$(\mathcal{L}^*)\begin{cases} \max & N(\sigma) \\ \text{s.t.} & \mathcal{R}(\mathcal{X};(\sigma_1,\cdots,\sigma_{L-1}))\neq\emptyset \\ & \sigma_k\in\{0,1\}^{n_k}, \forall k. \end{cases}$$

Proposition 1. *For each $k\in\{1,\cdots,L-1\}$ and each $i\in\{1,\ldots,n_k\}$, the function $g_k^i \mapsto N(g)$ is convex.*

Proof. The function is the composition of $g_k^i \mapsto M_L\mathrm{diag}(g_{L-1})M_{L-1}\cdots\mathrm{diag}(g_1)M_1$ which is affine and the norm $\|.\|$ which is convex, therefore it is convex [7, Proposition 2.1.5]. □

Proposition 2. *Problem $(\mathcal{U}^*)$ is equivalent to the following one:*

$$(\hat{\mathcal{U}})\begin{cases} \max & N(g) \\ \text{s.t.} & g_k^i \in \partial\mathrm{ReLU}(\theta_k^i(x)),\ \forall i\in\{1,\ldots,n_k\}, k\in\{1,\cdots,L-1\} \\ & g_k\in\{0,1\}^{n_k}, \forall k\in\{1,\cdots,L-1\} \\ & x\in\mathcal{X}. \end{cases}$$

Proof. Let g^* denote a solution of $(\mathcal{U}^*)$ and $\hat{g}$ a solution of $(\hat{\mathcal{U}})$.

The vector $\hat{g}$ is naturally feasible for $(\mathcal{U}^*)$ and thus

$$N(\hat{g}) \leq N(g^*).$$

Now, if $(g^*)_k^i \in]0,1[$ for some i and k, using convexity of $g_k^i \mapsto N(g)$, there exists $\overline{g}$, whose all components are equal to that of g^* except $(\overline{g})_k^i$ which belongs to $\{0,1\}$, such that $N(\overline{g}) \geq N(g^*)$. Moreover $(g^*)_k^i \in]0,1[\cap \partial\mathrm{ReLU}(\theta_k^i(x))$ implies that $\theta_k^i(x) = 0$ and thus $(\overline{g})_k^i \in \{0,1\}$ is also in $\partial\mathrm{ReLU}(\theta_k^i(x))$. Repeating this for all components of g^* that are in $]0,1[$, we show that there exists a solution $\overline{g}$ of $(\mathcal{U}^*)$ whose components are all in $\{0,1\}$ satisfying $N(g^*) \leq N(\overline{g})$. Therefore, we have

$$N(g^*) \leq N(\overline{g}) \leq N(\hat{g}).$$

□

Using Proposition 2 and the fact that $g_k^i \in \partial\mathrm{ReLU}(\theta_k^i(x))$ is equivalent to $\left(g_k^i - \frac{1}{2}\right)\theta_k^i(x) \geq 0$ for $g_k^i \in \{0,1\}$, we compute an upper bound of the Lipschitz constant by solving

$$(\overline{\mathcal{P}}) \begin{cases} \max & N(g) \\ \text{s.t.} & (g_k^i - \frac{1}{2})\theta_k^i(x) \geq 0,\ \forall i \in \{1,\dots,n_k\}, k \in \{1,\cdots,L-1\} \\ & g_k \in \{0,1\}^{n_k}, \forall k \in \{1,\cdots,L-1\} \\ & x \in \mathcal{X}. \end{cases}$$

By the definition of the activation region, problem $(\mathcal{L}^*)$ can be reformulated as

$$(\underline{\mathcal{P}}) \begin{cases} \max & N(\sigma) \\ \text{s.t.} & (\sigma_k^i - \frac{1}{2})\theta_k^i(x) > 0,\ \forall i \in \{1,\dots,n_k\}, k \in \{1,\cdots,L-1\} \\ & \sigma_k \in \{0,1\}^{n_k}, \forall k \in \{1,\cdots,L-1\} \\ & x \in \mathcal{X}. \end{cases}$$

In order to avoid strict inequalities, we introduce for $\varepsilon \geq 0$,

$$(\underline{\mathcal{P}})_\varepsilon \begin{cases} \max & N(\sigma) \\ \text{s.t.} & \left(\sigma_k^i - \frac{1}{2}\right)\theta_k^i(x) \geq \varepsilon,\ \forall i \in \{1,\dots,n_k\}, k \in \{1,\cdots,L-1\} \\ & \sigma_k \in \{0,1\}^{n_k}, \forall k \in \{1,\cdots,L-1\} \\ & x \in \mathcal{X}. \end{cases}$$

Observe that $(\underline{\mathcal{P}})_0$ corrresponds to $(\overline{\mathcal{P}})$.

We now introduce the following constraint sets:

$$\mathcal{C} = \mathcal{X} \times \{0,1\}^{n_1+n_2+\ldots+n_{L-1}}, \tag{4}$$

$$\mathcal{C}_s = \left\{(x,\sigma) \in \mathcal{C} : \left(\sigma_k^i - \frac{1}{2}\right)\theta_k^i(x) > 0,\ \forall i \in \{1,\dots,n_k\}, k \in \{1,\cdots,L-1\}\right\}, \tag{5}$$

$$\mathcal{C}_\varepsilon = \left\{(x,\sigma) \in \mathcal{C} : \left(\sigma_k^i - \frac{1}{2}\right)\theta_k^i(x) \geq \varepsilon,\ \forall i \in \{1,\dots,n_k\}, k \in \{1,\cdots,L-1\}\right\}, \tag{6}$$

so that the preceeding problems can be rewritten as:

$$(\overline{\mathcal{P}}) \left\{ \begin{array}{ll} \max_{x,\sigma} & N(\sigma) \\ \text{s.t.} & (x,\sigma) \in \mathcal{C}_0 \end{array} \right\} \qquad (\underline{\mathcal{P}}) \left\{ \begin{array}{ll} \max_{x,\sigma} & N(\sigma) \\ \text{s.t.} & (x,\sigma) \in \mathcal{C}_s \end{array} \right\} \qquad (\underline{\mathcal{P}})_\varepsilon \left\{ \begin{array}{ll} \max_{x,\sigma} & N(\sigma) \\ \text{s.t.} & (x,\sigma) \in \mathcal{C}_\varepsilon \end{array} \right\}.$$

Let $\overline{L}$, $\underline{L}$, $\underline{L}_\varepsilon$ denote the optimal values of $(\overline{\mathcal{P}})$, $(\underline{\mathcal{P}})$ and $(\underline{\mathcal{P}})_\varepsilon$ respectively. The following proposition summarizes and completes the above discussion.

Proposition 3. *We have*

1. *The function* $]0,+\infty[\ni \varepsilon \mapsto \underline{L}_\varepsilon$ *is non-increasing and piece-wise constant.*
2. $\lim_{\varepsilon\downarrow 0} \underline{L}_\varepsilon \uparrow \underline{L} \leq L(f,\mathcal{X}) \leq \overline{L} = \underline{L}_0$.
3. *If* $\cup_{i,k} H_{i,k}$ *is Lebesgue measure negligible, then* $L(f,\mathcal{X}) = \underline{L}$.
4. *If the* ReLU *network* f *is in general position (see Definition 4 in [9]), then* $L(f,\mathcal{X}) = \overline{L} = \underline{L}$.

Proof. 1. If $\varepsilon_1 \leq \varepsilon_2$ then $\mathcal{C}_{\varepsilon_2} \subset \mathcal{C}_{\varepsilon_1}$ which implies that $\underline{L}_\varepsilon$ is non increasing. Moreover, $\underline{L}_\varepsilon$ belongs to $\{N(\sigma), \sigma \in \{0,1\}^{n_1+n_2+\ldots+n_{L-1}}\}$ which is a finite set. Hence $\underline{L}_\varepsilon$ is piece-wise constant.

2. Observe that $\underline{L}_\varepsilon \leq \underline{L} \leq L(f,\mathcal{X}) \leq \overline{L} = \underline{L}_0$. To prove the remaining statement, let $\hat{\sigma}$ an optimal solution of $(\underline{\mathcal{P}})$. Then $\mathcal{R}(\mathcal{X};\hat{\sigma}) \neq \emptyset$, that is there exists $\hat{x} \in \mathcal{X}$ such that $\left(\hat{\sigma}_k^i - \frac{1}{2}\right)\theta_k^i(\hat{x}) > 0,\ \forall i \in \{1,\ldots,n_k\}, \forall k \in \{1,\cdots,L-1\}$. Define

$$\hat{\varepsilon} = \min_{i\in\{1,\ldots,n_k\}, k\in\{1,\cdots,L-1\}} \left(\hat{\sigma}_k^i - \frac{1}{2}\right)\theta_k^i(\hat{x}).$$

So $(\hat{x},\hat{\sigma}) \in \mathcal{C}_\varepsilon$ for all $\varepsilon \leq \hat{\varepsilon}$ and thus $\underline{L}_\varepsilon \geq \underline{L}$.

3. We can write $\mathcal{X} = (\bigcup_\sigma \mathcal{R}(\mathcal{X};\sigma)) \cup (\bigcup_{i,k} H_{i,k} \cap \mathcal{X})$. Now (1) implies

$$\begin{aligned} L(f,\mathcal{X}) &= \operatorname{esssup}_{x\in(\bigcup_\sigma \mathcal{R}(\mathcal{X};\sigma))} \sup_{G\in\mathcal{J}f(x)} \|G\| \\ &= \sup_{\sigma|\mathcal{R}(\mathcal{X},\sigma)\neq\emptyset} N(\sigma) \\ &= \underline{L}. \end{aligned}$$

4. By [9, Theorem 2] if the ReLU network is in general position then

$$\mathcal{J}f(x) = \left\{\|M_L \operatorname{diag}(g_{L-1}) M_{L-1} \cdots \operatorname{diag}(g_1) M_1\| \quad | \quad g_k \in [0,1]^{n_k},\ g_k^i \in \partial\mathrm{ReLU}(\theta_k^i(x))\right\}.$$

By (1) we obtain $L(f,\mathcal{X}) = \overline{L}$. Furthermore, if the ReLU network f is in a general position then $\cup_{i,k} H_{i,k}$ is Lebesgue measure negligible [9], which ensures the second equality using the preceding item.

□

Remark 1. *By [9, Theorem 3] the set of* ReLU *networks not in general position has Lebesgue measure zero over the parameter space and consequently for almost all* ReLU *Networks we have* $L(f,\mathcal{X}) = \overline{L} = \underline{L}$.

Corollary 1. *If* $\mathcal{X} = \mathbb{R}^{n_0}$ *and the bias* b_k, $k \in \{1,\cdots,L-1\}$ *are zero, then* $\underline{L}_\varepsilon = \underline{L}$ *for all* $\varepsilon > 0$.

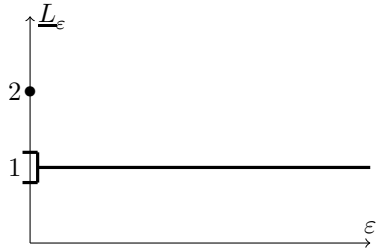
Proof. We have already established that $\underline{L}_\varepsilon \leq \underline{L}$. Let $(\hat{x},\hat{\sigma}) \in \mathcal{C}_s$ an optimal solution of $(\underline{\mathcal{P}})$ and let $\varepsilon > 0$. Since the bias b_k, $k \in \{1,\cdots,L-1\}$ are zero then for any $\lambda > 0$ we have $\theta_k(\lambda\hat{x}) = \lambda\theta_k(\hat{x})$. Therefore we can choose λ sufficiently large so that $\left(\hat{\sigma}_k^i - \frac{1}{2}\right)\theta_k^i(\lambda\hat{x}) > \varepsilon,\ \forall i \in \{1,\ldots,n_k\}, k \in \{1,\cdots,L-1\}$ ensuring that $(\lambda\hat{x},\hat{\sigma}) \in \mathcal{C}_\varepsilon$ and consequently $\underline{L}_\varepsilon \geq N(\hat{\sigma}) = \underline{L}$. □

The following simple examples illustrate the above results.

Exemple 1. *For* $f(x) = x - 1$, *we trivially have* $L(f,\mathcal{X}) = 1$. *This function can also be written* $f(x) = \max(x-1,0) - \max(1-x,0)$, *corresponding to our formalism with*

$$M_1 = \begin{pmatrix} 1 \\ -1 \end{pmatrix},\ M_2 = \begin{pmatrix} 1 & -1 \end{pmatrix},\ b_1 = \begin{pmatrix} -1 \\ 1 \end{pmatrix},\ b_2 = 0.$$

With this choice, we have $\underline{L}_\varepsilon = \underline{L} = 1$ *for all* $\varepsilon > 0$ *and* $\underline{L}_0 = \overline{L} = 2$. *Indeed,* $x = 1$ *is neither in* $\mathcal{R}(\mathbb{R};\sigma)$ *nor in the feasible set of* $(\underline{\mathcal{P}})_\varepsilon$ *for* $\varepsilon > 0$, *but it belongs to* $\mathcal{C}_0$.



Exemple 2. *For* $f(x) = \max(x+1,0) - \max(x-1,0)$ *corresponding to*

$$M_1 = \begin{pmatrix} 1 \\ 1 \end{pmatrix},\ M_2 = \begin{pmatrix} -1 & 1 \end{pmatrix},\ b_1 = \begin{pmatrix} -1 \\ 1 \end{pmatrix},\ b_2 = 0,$$

we have $\underline{L} = 1$, $L(f,\mathcal{X}) = 1$ *and* $\underline{L}_0 = \overline{L} = 1$. *Moreover,*

$$\underline{L}_\varepsilon = \begin{cases} 1 & \text{for all } 0 < \varepsilon \leq 0.5, \\ 0 & \text{for all } \varepsilon > 0.5. \end{cases}$$

4 MIQCQP reformulations

Let now consider the L^p-norm for $p \in \{1, 2, \infty\}$ and assume that $\mathcal{X}$ can be expressed by quadratic constraints (e.g. a ball) or linear constraints (e.g. polyhedron).

Then let us remark that we have $\rho_k \circ \theta_k(x) = \sigma_k \odot \theta_k(x)$ for all $k = 1, 2, \cdots L-1$ as soon as (x,σ) is in one of the sets $\mathcal{C}_\varepsilon$, $\mathcal{C}_s$ or $\mathcal{C}_0$. By definition of $\theta_k(x)$, we have thus $\rho_k \circ \theta_k(x) = \sigma_k \odot \theta_k(x) = \sigma_k \odot (M_k \circ \rho_{k-1} \circ \theta_{k-1}(x) + b_k)$ for $k = 2, 3, \cdots L-1$ and $\rho_1 \circ \theta_1(x) = \sigma_1 \odot (M_1 x + b_1)$. Denoting $x_k = \rho_k \circ \theta_k(x)$ we get the following bilinear relations

$$x_k = \sigma_k \odot (M_k x_{k-1} + b_k) \text{ for } k = 1, 2, \cdots L-1 \text{ with } x_0 = x.$$

Therefore, the constraints of $(\underline{\mathcal{P}})_\varepsilon$ are equivalent to

$$x_0 \in \mathcal{X} \text{ and for all } k \in \{1, 2, \ldots, L-1\}, \begin{cases} x_k = \sigma_k \odot (M_k x_{k-1} + b_k), \\ (\sigma_k - \frac{1}{2}) \odot (M_k x_{k-1} + b_k) \geq \varepsilon, \\ \sigma_k \in \{0,1\}^{n_k}. \end{cases}$$

The objective function of $(\underline{\mathcal{P}})_\varepsilon$ can also be expressed, depending on the chosen L^p-norm, as

$$N_p(\sigma) = \|M_L \mathrm{diag}(\sigma_{L-1}) M_{L-1} \cdots \mathrm{diag}(\sigma_1) M_1\|_p = \sup_{y, \|y\|_p \leq 1} \|M_L \mathrm{diag}(\sigma_{L-1}) M_{L-1} \cdots \mathrm{diag}(\sigma_1) M_1 y\|_p$$

and equivalently as

$$\begin{array}{ll}\max\limits_{y} & \|y_L\|_p \\ \text{s.t.} & y_k = M_k \text{diag}(\sigma_{k-1}) y_{k-1}, \forall k \in \{2,\dots,L\}, \\ & y_1 = M_1 y_0 \\ & \|y_0\|_p \leq 1\end{array}$$

where y is the collection of $y_0, y_1, \dots, y_L$.

Finally, for a given $p \in \{1, 2, \infty\}$, $(\underline{\mathcal{P}})_\varepsilon$ is equivalent to the following problem:

$$(\underline{\mathcal{P}}_p)_\varepsilon \begin{cases} \max\limits_{y,\sigma,x} & \|y_L\|_p \\ \text{s.t.} & y_k = M_k \text{diag}(\sigma_{k-1}) y_{k-1}, \forall k \in \{2,\dots,L\}, \\ & y_1 = M_1 y_0 \\ & x_k = \sigma_k \odot (M_k x_{k-1} + b_k), k = 1, 2, \cdots L-1 \\ & (\sigma_k - \frac{1}{2}) \odot (M_k x_{k-1} + b_k) \geq \varepsilon, \ k = 1, 2, \cdots L-1 \\ & \sigma_k \in \{0,1\}^{n_k}, k = 1, 2, \cdots L-1 \\ & x_0 \in \mathcal{X} \\ & \|y_0\|_p \leq 1 \end{cases}$$

where σ is the collection of $\sigma_1, \dots, \sigma_{L-1}$ and x is the collection of $x_0, x_1, \dots, x_{L-1}$. Provided that $\mathcal{X}$ is expressible by quadratic (or linear) constraints, the above problem can also be expressed as a MIQCQP as shown below.

The case $p = 2$ In this case, $(\underline{\mathcal{P}}_2)_\varepsilon$ is trivially equivalent to the following MIQCQP problem.

$$\begin{array}{ll}\max\limits_{y,\sigma,x} & \|y_L\|_2^2 \\ \text{s.t.} & y_k = M_k \text{diag}(\sigma_{k-1}) y_{k-1}, \forall k \in \{2,\dots,L\}, \\ & y_1 = M_1 y_0 \\ & x_k = \sigma_k \odot (M_k x_{k-1} + b_k), k = 1, 2, \cdots L-1 \\ & (\sigma_k - \frac{1}{2}) \odot (M_k x_{k-1} + b_k) \geq \varepsilon, \ k = 1, 2, \cdots L-1 \\ & \sigma_k \in \{0,1\}^{n_k}, k = 1, 2, \cdots L-1 \\ & x_0 \in \mathcal{X} \\ & \|y_0\|_2^2 \leq 1. \end{array}$$

The case $p = 1$ Observing that for all $a \in \mathbb{R}$, $|a| = (2\lambda - 1)a$ with $\lambda \in \{0, 1\}$ and $(2\lambda - 1)a \geq 0$, $(\underline{\mathcal{P}}_1)_\varepsilon$ is equivalent to

$$\begin{array}{ll}\max\limits_{y,\sigma,x,\nu,\mu} & \sum_{i=1}^{n_L} (2\mu_i - 1)(y_L)_i \\ \text{s.t.} & y_k = M_k \text{diag}(\sigma_{k-1}) y_{k-1}, \forall k \in \{2,\dots,L\}, \\ & y_1 = M_1 y_0 \\ & x_k = \sigma_k \odot (M_k x_{k-1} + b_k), k = 1, 2, \cdots L-1 \\ & (\sigma_k - \frac{1}{2}) \odot (M_k x_{k-1} + b_k) \geq \varepsilon, \ k = 1, 2, \cdots L-1 \\ & \sigma_k \in \{0,1\}^{n_k}, k = 1, 2, \cdots L-1 \\ & x_0 \in \mathcal{X} \\ & \nu \in \{0,1\}^{n_0} \\ & (2\nu_i - 1)(y_0)_i \geq 0, \ i = 1, 2, \cdots, n_0 \\ & \sum_{i=1}^{n_0} (2\nu_i - 1)(y_0)_i \leq 1, i = 1, 2, \cdots, n_0 \\ & \mu \in \{0,1\}^{n_L} \\ & (2\mu_i - 1)(y_L)_i \geq 0, \ i = 1, 2, \cdots, n_L. \end{array}$$

The last constraints can be further linearized by introducing additional binary variables and using a *big-M* technique. Indeed, for all $x \in \mathbb{R}$, with $|x| \leq B$, $u = |x|$ if and only if there exists $\lambda \in \{0, 1\}$ such

that

$$\begin{cases} u \geq x \text{ and } u \geq -x, \\ x \leq B(1-\lambda) \text{ and } x \geq -B\lambda, \\ u \leq -x + 2B(1-\lambda) \text{ and } u \leq x + 2B\lambda. \end{cases}$$

The first constraints imply that $u \geq |x|$. The second ones ensure that if $x > 0$ (respectively $x < 0$) then $\lambda = 0$ (respectively $\lambda = 1$). The last constraints guarantee that $u \leq |x|$. Problem $(\underline{\mathcal{P}}_1)_\varepsilon$ is hence equivalent to

$$\begin{array}{ll} \max\limits_{y,\sigma,x,u,w,\nu,\mu} & \sum_{i=1}^{n_L} w_i \\ \text{s.t.} & y_k = M_k \text{diag}(\sigma_{k-1}) y_{k-1}, \forall k \in \{2, \ldots, L\}, \\ & y_1 = M_1 y_0 \\ & x_k = \sigma_k \odot (M_k x_{k-1} + b_k), k = 1, 2, \cdots L-1 \\ & (\sigma_k - \frac{1}{2}) \odot (M_k x_{k-1} + b_k) \geq \varepsilon, \ k = 1, 2, \cdots L-1 \\ & \sigma_k \in \{0,1\}^{n_k}, k = 1, 2, \cdots L-1 \\ & x_0 \in \mathcal{X} \\ & \\ & \nu \in \{0,1\}^{n_0} \\ & -1 \leq y_0 \leq 1 \\ & (y_0)_i \leq u_i, \ i = 1, 2, \cdots, n_0 \\ & -(y_0)_i \leq u_i, \ i = 1, 2, \cdots, n_0 \\ & (y_0)_i \leq 1 - \nu_i, i = 1, 2, \cdots, n_0 \\ & (y_0)_i \geq -\nu_i, \ i = 1, 2, \cdots, n_0 \\ & u_i \leq -(y_0)_i + 2(1-\nu_i), \ i = 1, 2, \cdots, n_0 \\ & u_i \leq (y_0)_i + 2\nu_i, i \ = 1, 2, \cdots, n_0 \\ & \sum_{i=1}^{n_0} u_i \leq 1 \\ & \\ & \mu \in \{0,1\}^{n_L} \\ & (y_L)_i \leq w_i, \ i = 1, 2, \cdots, n_L \\ & -(y_L)_i \leq w_i, \ i = 1, 2, \cdots, n_L \\ & (y_L)_i \leq C(1-\mu_i), i = 1, 2, \cdots, n_L \\ & (y_L)_i \geq -C\mu_i, \ i = 1, 2, \cdots, n_L \\ & w_i \leq -(y_L)_i + 2C(1-\mu_i), \ i = 1, 2, \cdots, n_L \\ & w_i \leq (y_L)_i + 2C\mu_i, i = 1, 2, \cdots, n_L \end{array}$$

where $C \gg 0$ is the so-called *big-M* constant.

The case $p = \infty$ Problem $(\underline{\mathcal{P}}_\infty)_\varepsilon$ can similarly be expressed as

$$\begin{array}{ll} \max\limits_{y,\sigma,x,u,\mu,\eta} & \sum_{i=1}^{n_L} \eta_i u_i \\ \text{s.t.} & y_k = M_k \text{diag}(\sigma_{k-1}) y_{k-1}, \forall k \in \{2, \ldots, L\}, \\ & y_1 = M_1 y_0 \\ & x_k = \sigma_k \odot (M_k x_{k-1} + b_k), k = 1, 2, \cdots L-1 \\ & (\sigma_k - \frac{1}{2}) \odot (M_k x_{k-1} + b_k) \geq \varepsilon, \ k = 1, 2, \cdots L-1 \\ & \sigma_k \in \{0,1\}^{n_k}, k = 1, 2, \cdots L-1 \\ & x_0 \in \mathcal{X} \\ & -1 \leq y_0 \leq 1 \\ & \\ & \mu \in \{0,1\}^{n_L} \\ & u \geq 0 \\ & u_i = (2\mu_i - 1)(y_L)_i, \ i = 1, 2, \cdots, n_L \\ & \eta \in \{0,1\}^{n_L} \\ & \sum_{i=1}^{n_L} \eta_i = 1. \end{array}$$

Note: in this formulation the last constraints can also be linearized as in the case $p = 1$.

References

[1] Peter Bartlett, Dylan J. Foster, and Matus Telgarsky. Spectrally-normalized margin bounds for neural networks, 2017.

[2] Moustapha Cisse, Piotr Bojanowski, Edouard Grave, Yann Dauphin, and Nicolas Usunier. Parseval networks: improving robustness to adversarial examples. In *ICML'17: Proceedings of the 34th International Conference on Machine Learning*, page 854–863, 2017.

[3] Sourour Elloumi and Amélie Lambert. Global solution of non-convex quadratically constrained quadratic programs. *Optimization Methods and Software*, 34(1):98–114, 2019.

[4] Mahyar Fazlyab, Alexander Robey, Hamed Hassani, Manfred Morari, and George J. Pappas. Efficient and accurate estimation of lipschitz constants for deep neural networks, 2023.

[5] H. Gouk, E. Frank, B. Pfahringer, and M.J. Cree. Regularisation of neural networks by enforcing lipschitz continuity. *Machine Learning*, 110:393–416, 2021.

[6] Boris Hanin and David Rolnick. Deep relu networks have surprisingly few activation patterns. In H. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alché-Buc, E. Fox, and R. Garnett, editors, *Advances in Neural Information Processing Systems*, volume 32. Curran Associates, Inc., 2019.

[7] Jean-Baptiste Hiriart-Urruty and Claude Lemaréchal. *Convex analysis and minimization algorithms. I*, volume 305 of *Grundlehren der mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer-Verlag, Berlin, 1993. Fundamentals.

[8] Cyril Imbert. Support functions of clarke's generalized jacobian and of its plenary hull. *Nonlinear Analysis: Theory, Methods and Applications*, 29(8):1111–1125, 2002.

[9] Matt Jordan and Alexandros G Dimakis. Exactly computing the local lipschitz constant of relu networks. In H. Larochelle, M. Ranzato, R. Hadsell, M.F. Balcan, and H. Lin, editors, *Advances in Neural Information Processing Systems*, volume 33, pages 7344–7353. Curran Associates, Inc., 2020.

[10] Kevin Scaman and Aladin Virmaux. Lipschitz regularity of deep neural networks: analysis and efficient estimation. In *NeurIPS*, 2018.

[11] Yusuke Tsuzuku, Issei Sato, and Masashi Sugiyama. Lipschitz-margin training: Scalable certification of perturbation invariance for deep neural networks, 2018.