ANNALES
DE LA FACULTÉ
DES SCIENCES
DE TOULOUSE
Mathématiques

CAMILLO DE LELLIS, GUIDO DE PHILIPPIS, BERND KIRCHHEIM AND RICCARDO TIONE
Geometric measure theory and differential inclusions

Geometric measure theory and differential inclusions [(*)]

Camillo De Lellis [(1)], Guido De Philippis [(2)],
Bernd Kirchheim [(3)] and Riccardo Tione [(4)]

Abstract. — In this paper we consider Lipschitz graphs of functions which are stationary points of strictly polyconvex energies. Such graphs can be thought as integral currents, resp. varifolds, which are stationary for some elliptic integrands. The regularity theory for the latter is a widely open problem, in particular no counterpart of the classical Allard's theorem is known. We address the issue from the point of view of differential inclusions and we show that the relevant ones do not contain the class of laminates which are used in [23] and [26] to construct nonregular solutions. Our result is thus an indication that an Allard's type result might be valid for general elliptic integrands. We conclude the paper by listing a series of open questions concerning the regularity of stationary points for elliptic integrands.

Résumé. — Dans cet article on considère des graphes de fonctiones lipschitziennes qui sont points stationnaires d'énergies strictement polyconvexes. Ces graphes peuvent être pensés soit comme des courants entièrs soit comme des varifolds, qui sont stationnaires pour des intégrandes elliptiques. La théorie de la régularité pour ce genre d'intégrandes est un problème ouvert, en particulier il n'existe aucun version du classique théorème d'Allard. On étudie ce problème en adoptant le point de vue des inclusions différentielles, et on démontre que l'inclusion différentielle associée avec le problème de la stationnarité ne contient pas la classe des laminés utilisés en [23] et [26] pour construire des solutions qui ne sont pas régulières. Notre résultat suggère que un théorème de régularité à la Allard peut rester valide pour intégrandes elliptiques générales. On conclut ce travail avec des questions concernant le problème de régularité des points stationnaires des intégrandes elliptiques.

1. Introduction

Let $\Omega \subset \mathbb{R}^m$ be open and $f \in C^1(\mathbb{R}^{n \times m}, \mathbb{R})$ be a (strictly) polyconvex function, i.e. such that there is a (strictly) convex $g \in C^1$ such that $f(X) = g(\Phi(X))$, where $\Phi(X)$ denotes the vector of subdeterminants of X of all orders. We then consider the following *energy* $\mathbb{E} : \mathrm{Lip}(\Omega, \mathbb{R}^n) \to \mathbb{R}$:

$$\mathbb{E}(u) := \int_\Omega f(Du)\, \mathrm{d}x. \tag{1.1}$$

DEFINITION 1.1. — *Consider a map $\bar{u} \in \mathrm{Lip}(\Omega, \mathbb{R}^n)$. The one-parameter family of functions $\bar{u} + \varepsilon v$ will be called* outer variations *and $\bar{u}$ will be called* critical for $\mathbb{E}$ *if*

$$\left. \frac{\mathrm{d}}{\mathrm{d}\varepsilon} \right|_{\varepsilon=0} \mathbb{E}(\bar{u} + \varepsilon v) = 0 \qquad \forall\, v \in C_c^\infty(\Omega, \mathbb{R}^n).$$

Given a vector field $\Phi \in C_c^1(\Omega, \mathbb{R}^m)$ we let X_ε be its flow.[1] *The one-parameter family of functions $u_\varepsilon = \bar{u} \circ X_\varepsilon$ will be called an* inner variation. *A critical point $\bar{u} \in \mathrm{Lip}(\Omega, \mathbb{R}^n)$ is* stationary *for $\mathbb{E}$ if*

$$\left. \frac{\mathrm{d}}{\mathrm{d}\varepsilon} \right|_{\varepsilon=0} \mathbb{E}(u_\varepsilon) = 0 \qquad \forall\, \Phi \in C_c^1(\Omega, \mathbb{R}^m).$$

Classical computations reduce the two conditions above to,[2] respectively,

$$\int_\Omega \langle Df(D\bar{u}), Dv \rangle\, \mathrm{d}x = 0 \qquad \forall\, v \in C_c^1(\Omega, \mathbb{R}^n). \tag{1.2}$$

and

$$\int_\Omega \langle Df(D\bar{u}), D\bar{u}D\Phi \rangle\, \mathrm{d}x - \int_\Omega f(D\bar{u}) \operatorname{div} \Phi\, \mathrm{d}x = 0 \quad \forall\, \Phi \in C_c^1(\Omega, \mathbb{R}^m). \tag{1.3}$$

The graphs of Lipschitz functions can be naturally given the structure of integer rectifiable currents (without boundary in $\Omega \times \mathbb{R}^m$) and of integral varifold, cf. [14, 16, 25]. In particular the graph of any stationary point $\bar{u} \in \mathrm{Lip}(\Omega, \mathbb{R}^n)$ for a polyconvex energy $\mathbb{E}$ can be thought as a stationary point for a corresponding elliptic energy, in the space of integer rectifiable currents and in that of integral varifolds, respectively, see [17, Chapter 1, Section 2]. Even though this fact is probably well known, it is not entirely trivial and we have not been able to find a reference in the literature: we

---

[1] Namely $X_\varepsilon(x) = \gamma_x(\varepsilon)$, where γ_x is the solution of the ODE $\gamma'(t) = \Phi(\gamma(t))$ subject to the initial condition $\gamma(0) = x$.

[2] $\langle A, B \rangle := \operatorname{tr}(A^T B)$ denotes the usual Hilbert–Schmidt scalar product of the matrices A and B.

therefore give a proof for the reader's convenience. Note that a particular example of polyconvex energy is given by the area integrand

$$\mathcal{A}(X) = \sqrt{\det(\mathrm{id}_{\mathbb{R}^{m \times m}} + X^T X)}. \tag{1.4}$$

The latter is *strongly polyconvex* when restricted to the any ball $B_R \subset \mathbb{R}^{n \times m}$, namely there is a positive constant $\varepsilon(R)$ such that $X \mapsto \mathcal{A}(X) - \varepsilon(R)|X|^2$ is still polyconvex on B_R.

When $n = 1$ strong polyconvexity reduces to locally uniform convexity and any Lipschitz critical point is therefore $C^{1,\alpha}$ by the De Giorgi–Nash theorem. The same regularity statement holds in the much simpler "dual case" $m = 1$, where criticality implies that the vector valued map $\bar{u}$ satisfies an appropriate system of ODEs. Remarkably, L. Székelyhidi in [26] proved the existence of smooth strongly polyconvex integrands $f : \mathbb{R}^{2 \times 2} \to \mathbb{R}$ for which the corresponding energy has Lipschitz critical points which are nowhere C^1. The paper [26] is indeed an extension of a previous groundbreaking result of S. Müller and V. Šverák [23], where the authors constructed a Lipschitz critical point to a smooth strongly quasiconvex energy (cf. [23] for the relevant definition) which is nowhere C^1. A precursor of such examples can be found in the pioneering PhD thesis of V. Scheffer, [24].

Minimizers of strongly quasiconvex functions have been instead proved to be regular almost everywhere, see [12, 20, 24]. Note that the "geometric" counterpart of the latter statement is Almgren's celebrated regularity theorem for integral currents minimizing strongly elliptic integrands [5]. Let us remark that stationary points need not to be local minimizers for the energy. This is proved, for instance, in the case $f(X) = \mathcal{A}(X)$ for $n = m = 2$, by H. Lawson and R. Osserman in [21, Theorem 5.3]. Standard computations show, on the other hand, that every minimizer for an energy is a stationary point. Moreover, combining the uniqueness result in [29] and [23, Theorem 4.1], it is easy to see that there exist critical points that are not stationary.

Other than the result in [29], not much is known about the properties of stationary points, in particular it is not known whether they must be C^1 on a set of full measure. Observe that Allard's ε regularity theorem applies when f is the area integrand and allows to answer positively to the latter question for f in (1.4). The validity of an Allard-type ε-regularity theorem for general elliptic energies is however widely open. A first interesting question is whether one could extend the examples of Müller and Šveràk and Székelyhidi to provide counterexamples. Both in [23] and [26], the starting point of the construction of irregular solutions is rewriting the condition (1.2) as a differential inclusion, and then finding a so-called T_N-configuration ($N = 4$

in the first case, $N = 5$ in the latter) in the set defining the differential inclusion. The main result of the present paper shows that such strategy fails in the case of stationary points. More precisely:

(a) We show that $\bar{u}$ solves (1.2), (1.3) if and only if there exists an L^∞ matrix field A that solves a certain system of linear, constant coefficients, PDEs and takes almost everywhere values in a fixed set of matrices, which we denote by K_f and call *inclusion set*, cf. Lemma 2.2. The latter system of PDEs will be called a *div-curl* differential inclusion, in order to distinguish them from classical differential inclusions, which are PDE of type $Du \in K$ a.e., and from "divergence differential inclusions" as for instance considered in [8].

(b) We give the appropriate generalization of T_N configurations for *div-curl* differential inclusions, which we will call T'_N configurations, cf. Definition 2.6. As in the "classical" case the latter are subsets of cardinality N of the set K_f which satisfy a particular set of conditions.

(c) We then prove the following nonexistence result.

THEOREM 1.2. — *If $f \in C^1(\mathbb{R}^{n \times m})$ is strictly polyconvex, then K_f does not contain any set $\{A_1, \ldots, A_N\}$ which induces a T'_N configuration.*

Remark 1.3 (Székelyhidi's result). — Theorem 1.2 can be directly compared with the results in [26], which concern the "classical" differential inclusions induced by (1.2) alone. In particular [26, Theorem 1] shows the existence of a smooth strongly polyconvex integrand $f \in C^\infty(\mathbb{R}^{2 \times 2})$ for which the corresponding "classical" differential inclusion contains a T_5 configuration (cf. Definition 2.5). In fact the careful reader will notice that the 5 matrices given in [26, Example 1] are incorrect. This is due to an innocuous sign error of the author in copying their entries. While other T_5 configurations can be however easily computed following the approach given in [26], according to [28], the correct original ones of [26, Example 1] are the following:

$$Z_1 \doteq \begin{pmatrix} 2 & 2 \\ -2 & -2 \\ -20 & -20 \\ -14 & -14 \end{pmatrix}, \quad Z_2 \doteq \begin{pmatrix} 3 & 5 \\ -5 & -9 \\ 0 & 10 \\ -3 & 1 \end{pmatrix}, \quad Z_3 \doteq \begin{pmatrix} 4 & 3 \\ -9 & -5 \\ 41 & 0 \\ 21 & -3 \end{pmatrix},$$

$$Z_4 \doteq \begin{pmatrix} -3 & -3 \\ 8 & 9 \\ -54 & -72 \\ -30 & -41 \end{pmatrix}, \quad Z_5 \doteq \begin{pmatrix} 0 & 0 \\ -1 & -2 \\ 18 & 36 \\ 11 & 22 \end{pmatrix}.$$

These five matrices form a T_5 configuration with $k_i = 2, \forall\, 1 \leqslant i \leqslant 5$, $P = 0$, and rank-one "arms" given by

$$
C_1 \doteq \begin{pmatrix} 1 & 1 \\ -1 & -1 \\ -10 & -10 \\ -7 & -7 \end{pmatrix}, \quad
C_2 \doteq \begin{pmatrix} 1 & 2 \\ -2 & -4 \\ 5 & 10 \\ 2 & 4 \end{pmatrix}, \quad
C_3 \doteq \begin{pmatrix} 1 & 0 \\ -3 & 0 \\ 23 & 0 \\ 13 & 0 \end{pmatrix},
$$

$$
C_4 \doteq \begin{pmatrix} -3 & -3 \\ 7 & 7 \\ -36 & -36 \\ -19 & -19 \end{pmatrix}, \quad
C_5 \doteq \begin{pmatrix} 0 & 0 \\ -1 & -2 \\ 18 & 36 \\ 11 & 22 \end{pmatrix}.
$$

Even though it seems still early to conjecture the validity of partial regularity for stationary points, our result leans toward a positive conclusion: Theorem 1.2 can be thought as a first step in that direction.

Another indication that an Allard type ε-regularity theorem might be valid for at least some class of energies is provided by the recent paper [9] of A. De Rosa, the second named author and F. Ghiraldin, which generalizes Allard's rectifiability theorem to stationary varifolds of a wide class of energies. In fact the authors' theorem characterizes in terms of an appropriate condition on the integrand (called "atomic condition", cf. [9, Definition 1.1]) those energies for which rectifiability of stationary points hold. Furthermore one can use the ideas in [9] to show that the atomic conditions implies strong $W^{1,p}$ convergence of sequences of stationary equi-Lipschitz graphs, [11]. When transported to stationary Lipschitz graphs, the latter is yet another obstruction to applying the methods of [23] and [26]. Indeed, in order to apply convex integration methods, one exploits the non-triviality of a suitable subset of the convex hull of K_f, the rank-one convex hull. Strong convergence of sequences of equi-Lipschitz stationary graphs impose that this hull is trivial. We refer the interested reader to Questions 8.7 and 8.10 for more explanations. In [10] it has been shown that the atomic condition implies Almgren's ellipticity. It is an intriguing issue to understand if this implication can be reversed and (if not) to understand wether this (hence stronger) assumption on the integrand can be helpful in establishing regularity of stationary points.

We believe that the connection between differential inclusions and geometric measure theory might be fruitful and poses a number of interesting and challenging questions. We therefore conclude this work with some related problems in Section 8.

The rest of the paper is organized as follows: in Section 2 we rewrite the Euler Lagrange conditions (1.2) and (1.3) as a div-curl differential inclusion and we determine its wave cone. We then introduce the inclusion set K_f and,

after recalling the definition of T_N configurations for classical differential inclusions, we define corresponding T'_N configurations for div-curl differential inclusions. In Section 3 we give a small extension of a key result of [27] on classical T_N configurations. In Section 4 we consider arbitrary sets of N matrices and give an algebraic characterization of those sets which belong to an inclusion set K_f for some strictly polyconvex f. In Section 5 we then prove the main theorem of the paper, Theorem 1.2. As already mentioned, Section 6 discusses the link between stationary graphs and stationary varifolds, whereas Section 8 is a collection of open questions.

2. Div-curl differential inclusions, wave cones and inclusion sets

As written in the introduction, the Euler–Lagrange conditions for energies $\mathbb{E}$ are given by:

$$\begin{cases} \displaystyle\int_\Omega \langle Df(Du), Dv \rangle \, \mathrm{d}x = 0 & \forall \, v \in C_c^1(\Omega, \mathbb{R}^n) \\ \displaystyle\int_\Omega \langle Df(Du), DuD\Phi \rangle \, \mathrm{d}x - \int_\Omega f(Du) \operatorname{div} \Phi \, \mathrm{d}x = 0 & \forall \, \Phi \in C_c^1(\Omega, \mathbb{R}^m), \end{cases} \tag{2.1}$$

Here we rewrite the system (2.1) as a differential inclusion. To do so, it is sufficient to notice that the left hand side of the second equation can be rewritten as

$$\int_\Omega \langle Df(Du), DuD\Phi \rangle \, \mathrm{d}x - \int_\Omega f(Du) \operatorname{div} \Phi \, \mathrm{d}x$$
$$= \int_\Omega \langle Du^T Df(Du), D\Phi \rangle - \langle f(Du) \operatorname{id}, D\Phi \rangle \, \mathrm{d}x$$
$$= \int_\Omega \langle Du^T Df(Du) - f(Du) \operatorname{id}, D\Phi \rangle \, \mathrm{d}x$$

Hence, the inner variation equation is the weak formulation of

$$\operatorname{div}(Du^T Df(Du) - f(Du) \operatorname{id}) = 0.$$

Since also the outer variation is the weak formulation of a PDE in divergence form, namely

$$\operatorname{div}(Df(Du)) = 0,$$

we introduce the following terminology:

DEFINITION 2.1. — *A div-curl differential inclusion is the following system of partial diffential equations for a triple of maps $X, Y \in L^\infty(\Omega, \mathbb{R}^{n \times m})$ and $Z \in L^\infty(\Omega, \mathbb{R}^{m \times m})$:*

$$\operatorname{curl} X = 0, \qquad \operatorname{div} Y = 0, \qquad \operatorname{div} Z = 0, \tag{2.2}$$

$$W := \begin{pmatrix} X \\ Y \\ Z \end{pmatrix}$$

$$\in K_f := \left\{ A \in \mathbb{R}^{(2n+m)\times m} : A = \begin{pmatrix} X \\ Df(X) \\ X^T Df(X) - f(X)\,\mathrm{id} \end{pmatrix} \right\}, \quad (2.3)$$

where $f \in C^1(\mathbb{R}^{n\times m})$ is a fixed function. The subset $K_f \subset \mathbb{R}^{(2n+m)\times m}$ will be called the inclusion set relative to f.

The following lemma is then an obvious consequence of the above discussion

LEMMA 2.2. — Let $f \in C^1(\mathbb{R}^{n\times m})$. A map $u \in \mathrm{Lip}(\Omega, \mathbb{R}^n)$ is a stationary point of the energy (1.1) if and only there are matrix fields $Y \in L^\infty(\Omega, \mathbb{R}^{n\times m})$ and $Z \in L^\infty(\Omega, \mathbb{R}^{m\times m})$ such that $W = (Du, Y, Z)$ solves the div-curl differential inclusion (2.2) and (2.3).

2.1. Wave cone for div-curl differential inclusions

We recall here the definition of wave cone for a system of linear constant coefficient first order PDEs. Given a system of linear constant coefficients PDEs

$$\sum_{i=1}^m A_i \partial_i z = 0 \qquad (2.4)$$

in the unknown $z : \mathbb{R}^m \supset \Omega \to \mathbb{R}^d$ we consider plane wave solutions to (2.4), that is, solutions of the form

$$z(x) = a h(x \cdot \xi), \qquad (2.5)$$

where $h : \mathbb{R} \to \mathbb{R}$. The wave cone Λ is given by the states $a \in \mathbb{R}^d$ for which there is a $\xi \neq 0$ such that for any choice of the profile h the function (2.5) solves (2.4), that is,

$$\Lambda := \left\{ a \in \mathbb{R}^d : \exists\, \xi \in \mathbb{R}^m \setminus \{0\} \ \text{ with } \ \sum_{i=1}^m \xi_i A_i a = 0 \right\}. \qquad (2.6)$$

The following lemma is then an obvious consequence of the definition and its proof is left to the reader.

LEMMA 2.3. — The wave cone of the system $\mathrm{curl}\, X = 0$ is given by rank one matrices, whereas the wave cone for the system (2.2) is given by triple of matrices (X, Y, Z) for which there is a unit vector $\xi \in \mathbb{S}^{m-1}$ and a vector $u \in \mathbb{R}^n$ such that $X = u \otimes \xi$, $Y\xi = 0$ and $Z\xi = 0$.

Motivated by the above lemma we then define

DEFINITION 2.4. — *The cone $\Lambda_{dc} \subset \mathbb{R}^{(2n+m)\times m}$ consists of the matrices in block form*

$$\begin{pmatrix} X \\ Y \\ Z \end{pmatrix}$$

with the property that there is a direction $\xi \in \mathbb{S}^{m-1}$ and a vector $u \in \mathbb{R}^n$ such that $X = u \otimes \xi$, $Y\xi = 0$ and $Z\xi = 0$.

2.2. T_N configurations

We start defining T_N configurations for "classical" differential inclusions.

DEFINITION 2.5. — *An ordered set of $N \geqslant 2$ matrices $\{X_i\}_{i=1}^N \subset \mathbb{R}^{n\times m}$ of distinct matrices is said to* induce *a T_N configuration if there exist matrices $P, C_i \in \mathbb{R}^{n\times m}$ and real numbers $k_i > 1$ such that:*

(a) *Each C_i belongs to the wave cone of $\operatorname{curl} X = 0$, namely $\operatorname{rank}(C_i) \leqslant 1$ for each i;*

(b) *$\sum_i C_i = 0$;*

(c) *$X_1, \ldots, X_n$, P and $C_1, \ldots, C_N$ satisfy the following N linear conditions*

$$\begin{aligned} X_1 &= P + k_1 C_1, \\ X_2 &= P + C_1 + k_2 C_2, \\ &\;\;\vdots \\ X_N &= P + C_1 + \cdots + k_N C_N\,. \end{aligned} \tag{2.7}$$

In the rest of the note we will use the word T_N configuration for the data $P, C_1, \ldots, C_N, k_1, \ldots k_N$. We will moreover say that the configuration is non-degenerate *if $\operatorname{rank}(C_i) = 1$ for every i.*

Note that our definition is more general that the one usually given in the literature (cf. [23, 26, 27]) because we drop the requirement that there are no rank one connections between distinct X_i and X_j. Moreover, rather than calling $\{X_1, \ldots, X_N\}$ a T_N configuration, we prefer to say that it "induces" a T_N configuration, namely we regard the whole data $X_1, \ldots, X_N, C_1, \ldots, C_N$, $k_1, \ldots, k_N$ since it is not at all clear that given an ordered set $\{X_1, \ldots, X_N\}$ of distinct matrices there is at most one choice of the matrices $C_1, \ldots, C_N$ and of the coefficients $k_1, \ldots, k_N$ satisfying the conditions above (if we drop
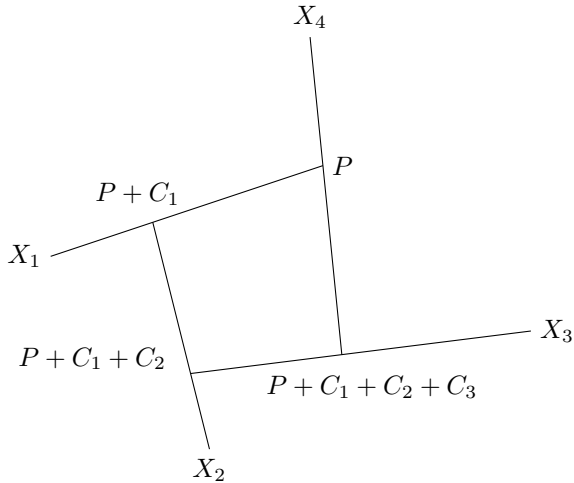
Figure 2.1. The geometric arrangement of a T_4 configuration.

the condition that the set is ordered, then it is known that there is more than one choice, see [15]).

We observe that the definition of T_N configuration could be split into two parts. A "geometric part", namely the points (b) and (c), can be considered as characterizing a certain "arrangement of $2N$ points" in the space of matrices, consisting of:

- A closed piecewise linear loop, loosely speaking a polygon (not necessarily *planar*) with vertices $P_1 = P + C_1$, $P_2 = P + C_1 + C_2$, ..., $P_N = P + C_1 + \cdots + C_N = P$;
- N additional "arms" which extend the sides of the polygon, ending in the points $X_1, \ldots, X_N$.

See Figure 2.1 for a graphical illustration of these facts in the case $N = 4$.

The closing condition in Definition 2.5(b) is a necessary and sufficient condition for the polygonal line to "close". Condition (c) determines that each X_i is a point on the line containing the segment $P_{i-1}P_i$. Note that the inequality $k_i > 1$ ensures that X_i is external to the segment, "on the side of P_i". The "nondegeneracy" condition is equivalent to the vertices of the polygon being all distinct. Note moreover that, in view of our definition, we include the possibility $N = 2$. In the latter case the T_2 configuration consists of a single rank one line and of 4 points X_1, X_2, C_1, C_2 lying on it.

We have decided to follow this convention, even though this is an unusual choice compared to the literature.

The second part of the Definition, namely condition (a), is of algebraic nature and related to the fact that T_N configurations are used to study "classical differential inclusions", namely PDEs of the form $\operatorname{curl} X = 0$. The condition prescribes simply that each vector $X_i - P_i$ belongs to the wave cone of $\operatorname{curl} X = 0$.

2.3. T'_N configurations

In this section we generalize the notion of T_N configuration to div-curl differential inclusions. The geometric arrangement remains the same, while the wave cone condition is replaced by the one dictated by the new PDE (2.2).

DEFINITION 2.6. — *A family* $\{A_1, \ldots, A_N\} \subset \mathbb{R}^{(2n+m) \times m}$ *of* $N \geqslant 2$ *distinct*

$$A_i := \begin{pmatrix} X_i \\ Y_i \\ Z_i \end{pmatrix}$$

induces a T'_N *configuration if there are matrices* $P, Q, C_i, D_i \in \mathbb{R}^{n \times m}$, $R, E_i \in \mathbb{R}^{m \times m}$ *and coefficients* $k_i > 1$ *such that*

$$\begin{pmatrix} X_i \\ Y_i \\ Z_i \end{pmatrix} = \begin{pmatrix} P \\ Q \\ R \end{pmatrix} + \begin{pmatrix} C_1 \\ D_1 \\ E_1 \end{pmatrix} + \cdots + \begin{pmatrix} C_{i-1} \\ D_{i-1} \\ E_{i-1} \end{pmatrix} + k_i \begin{pmatrix} C_i \\ D_i \\ E_i \end{pmatrix} \quad (2.8)$$

for every i, and the following properties hold:

(a) *each element* (C_i, D_i, E_i) *belongs to the wave cone* Λ_{dc} *of* (2.2);
(b) $\sum_\ell C_\ell = 0$, $\sum_\ell D_\ell = 0$ *and* $\sum_\ell E_\ell = 0$.

We say that the T'_N *configuration is* nondegeneate *if* $\operatorname{rank}(C_i) = 1$ *for every i.*

We collect here some simple consequences of the definition above and of the discussion on T_N configurations.

PROPOSITION 2.7. — *Assume* $A_1, \ldots, A_N$ *induce a* T'_N *configuration with* P, Q, R, C_i, D_i, E_i *and* k_i *as in Definition 2.6. Then:*

(i) $\{X_1, \ldots, X_N\}$ *induce a* T_N *configuration of the form* (2.7), *if they are distinct; moreover the* T'_N *configuration is nondegenerate if and only if the* T_N *configuration induced by* $\{X_1, \ldots, X_N\}$ *is nondegenerate;*

(ii) *For each i there is an $n_i \in \mathbb{S}^{m-1}$ and a $u_i \in \mathbb{R}^n$ such that $C_i = u_i \otimes n_i$, $D_i n_i = 0$ and $E_i n_i = 0$;*

(iii) $\operatorname{tr} C_i^T D_i = \langle C_i, D_i \rangle = 0$ *for every i.*

Proof. — (i) and (ii) are an obvious consequence of Definition 2.6 and of Definition 2.4. After extending n_i to an orthonormal basis $\{n_i, v_2^j, \dots v_m^j\}$ of $\mathbb{R}^m$ we can explicitely compute

$$\langle C_i, D_i \rangle = (D_i n_i, C_i n_i) + \sum_{j=2}^{m} (D_i v_i^j, C_i v_i^j) = 0\,,$$

where $(\,\cdot\,,\cdot\,)$ denotes the Euclidean scalar product. □

2.4. Strategy

Before starting with the proof of the main result of this part of the paper, it is convenient to explain the strategy we intend to follow. In order to do so, let us consider the simplest case $n = m = 2$, $N = 5$. Suppose by contradiction that there exists a strictly polyconvex function $f : \mathbb{R}^{2\times2} \to \mathbb{R}$, $f(X) = g(X, \det(X))$ and a T_5' configuration A_1, A_2, A_3, A_4, A_5,

$$A_i = \begin{pmatrix} X_i \\ Y_i \\ Z_i \end{pmatrix}, \quad \forall\, i \in \{1, \dots, 5\},$$

where X_i, Y_i, Z_i fulfill the relations of (2.8), i.e.

$$\begin{pmatrix} X_i \\ Y_i \\ Z_i \end{pmatrix} = \begin{pmatrix} P \\ Q \\ R \end{pmatrix} + \begin{pmatrix} C_1 \\ D_1 \\ E_1 \end{pmatrix} + \cdots + \begin{pmatrix} C_{i-1} \\ D_{i-1} \\ E_{i-1} \end{pmatrix} + k_i \begin{pmatrix} C_i \\ D_i \\ E_i \end{pmatrix}.$$

For convenience, let us consider $P = Q = 0, \operatorname{tr}(R) = 0$. We will prove in Lemma 5.1 that this can be done without loss of generality. It is convenient to think of the relations $A_i \in K_f$, $\forall\, i$ as two separate pieces of information:

$$\begin{pmatrix} X_i \\ Y_i \end{pmatrix} \in K_f' = \left\{ A \in \mathbb{R}^{4\times2} : A = \begin{pmatrix} X \\ Df(X) \end{pmatrix}, X \in \mathbb{R}^{2\times2} \right\} \qquad (2.9)$$

and

$$Z_i = X_i^T Y_i - f(X_i)\,\mathrm{id}\,. \qquad (2.10)$$

Let us denote with $c_i := f(X_i)$. Similarly to the procedure of [26], we exploit the polyconvexity of f to rewrite (2.9) in terms of inequalities involving X_i, Y_i, c_i, d_i, where $d_i := \partial_{y_5} g(y_1, y_2, y_3, y_4, y_5)\big|_{(X_i, \det(X_i))}$, of the form

$$c_i - c_j + \langle Y_i, X_j - X_i \rangle - d_i \det(X_i - X_j) < 0. \qquad (2.11)$$

This is the content of Proposition 4.1. The final goal is to prove that these inequalities can not be fulfilled at the same time. The previous expression

can be considerably simplified by the structure result on T_N configurations in $\mathbb{R}^{2\times 2}$ of [27, Proposition 1]. This asserts, in the specific case of the ongoing example, the existence of 5 vectors $(t_1^i, \ldots, t_5^i), i \in \{1, \ldots, 5\}$ with positive components, such that

$$\sum_{j=1}^{5} t_j^i \det(X_j - X_i) = 0. \tag{2.12}$$

If we use this result in (2.11), we can eliminate from the expression the "variable" d_i, thus obtaining

$$\nu_i := \sum_{j=1}^{5} t_j^i (c_i - c_j + \langle Y_i, X_j - X_i \rangle - d_i \det(X_i - X_j))$$

$$= \sum_{j=1}^{5} t_j^i (c_i - c_j + \langle Y_i, X_j - X_i \rangle) < 0, \quad \forall\, i \in \{1, \ldots, 5\},$$

compare Corollary 4.3. Section 3 is devoted to extending relations (2.12) to general T_N configurations in $\mathbb{R}^{n\times m}$. Despite being very useful, this simplification can not conclude the proof. Indeed, up to now we have exploited (2.9) and the fact that $\{X_1, \ldots, X_5\}$ induce a T_5 configuration, but this is the exact same situation of [26]. Since from that paper we know the existence of T_5 configurations in K_f', clearly we can not reach a contradiction at this point of the strategy. This is where the inner variations come into play. In the proof of Theorem 1.2, we rewrite (2.10) using the definition of T_5' configuration and, after some manipulations, we find that ν_i are "generalized eigenvalues" of R, see (5.12), where R is the matrix appearing in the definition of A_i as T_N' configuration. Since we supposed $\operatorname{tr}(R) = 0$, we find a subset $\mathcal{I} \subset \{1, \ldots, 5\}$ such that

$$\sum_{i\in\mathcal{I}} \xi_i \nu_i = 0,$$

for positive coefficients ξ_i. This implies that ν_i can not be strictly negative for every i, therefore reaching the desired contradiction.

3. Preliminaries on classical T_N configurations

This section is devoted to a slight generalization of a powerful machinery introduced in [27] to study T_N configurations.

3.1. Székelyhidi's characterization of T_N configurations in $\mathbb{R}^{2\times 2}$

We start with the following elegant characterization.

PROPOSITION 3.1 ([27, Proposition 1]). — *Given a set $\{X_1,\ldots,X_N\} \subset \mathbb{R}^{2\times 2}$ and $\mu \in \mathbb{R}$, we let A^μ be the following $N \times N$ matrix:*

$$
A^\mu := \begin{bmatrix}
0 & \det(X_1 - X_2) & \det(X_1 - X_3) & \ldots & \det(X_1 - X_N) \\
\mu \det(X_1 - X_2) & 0 & \det(X_2 - X_3) & \ldots & \det(X_2 - X_N) \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
\mu \det(X_1 - X_N) & \mu \det(X_2 - X_N) & \mu \det(X_3 - X_N) & \ldots & 0
\end{bmatrix}.
$$

Then, $\{X_1,\ldots,X_N\}$ induces a T_N configuration if and only if there exists a vector $\lambda \in \mathbb{R}^N$ with positive components and $\mu > 1$ such that

$$
A^\mu \lambda = 0.
$$

Even though not explicitely stated in [27], the following Corollary is part of the proof of Proposition 3.1 and it is worth stating it here again, since we will make extensive use of it in the sequel.

COROLLARY 3.2. — *Let $\{X_1,\ldots,X_N\} \subset \mathbb{R}^{2\times 2}$ and let $\mu > 1$ and $\lambda \in \mathbb{R}^N$ be a vector with positive entries such that $A^\mu \lambda = 0$. Define the vectors*

$$
t^i := \frac{1}{\xi_i}(\mu\lambda_1,\ldots,\mu\lambda_{i-1},\lambda_i,\ldots,\lambda_N), \quad for \ i \in \{1,\ldots,N\} \tag{3.1}
$$

where $\xi_i > 0$ is a normalizing constant so that $\|t^i\|_1 := \sum_j |t^i_j| = 1, \forall \ i$. Define the matrices C_j with $j \in \{1,\ldots,N-1\}$ and P by solving recursively

$$
\sum_{j=1}^N t^i_j X_j = P + C_1 + \cdots + C_{i-1} \tag{3.2}
$$

and set $C_N := -C_1 - \cdots - C_{N-1}$. Finally, define

$$
k_i = \frac{\mu\lambda_1 + \cdots + \mu\lambda_i + \lambda_{i+1} \cdots + \lambda_N}{(\mu - 1)\lambda_i}. \tag{3.3}
$$

Then $P, C_1,\ldots,C_N$ and $k_1,\ldots k_N$ give a T_N configuration induced by $\{X_1,\ldots,X_N\}$ (i.e. (2.7) holds).

Moreover, the following relation holds for every i:

$$
\det\left(\sum_{j=1}^N t^i_j X_j\right) = \sum_{j=1}^N t^i_j \det(X_j). \tag{3.4}
$$

Remark 3.3. — Observe that the relations (3.3) can be inverted in order to compute μ and λ (the latter up to scalar multiples) in terms of $k_1, \ldots, k_N$. In fact, let us impose

$$\|\lambda\|_1 = \lambda_1 + \cdots + \lambda_N = 1 \,.$$

Then, regarding μ as a parameter, the equations (3.3) give a linear system in triangular form which can be explicitely solved recursively, giving the formula

$$\lambda_j = \frac{k_1 k_2 \cdots k_{j-1}}{(\mu - 1)(k_1 - 1)(k_2 - 1)\cdots(k_j - 1)} \,. \tag{3.5}$$

The following identity can easily be proved by induction:

$$\frac{1}{k_1 - 1} + \frac{k_1}{(k_1 - 1)(k_2 - 1)} + \cdots + \frac{k_1 \cdots k_{j-1}}{(k_1 - 1)\cdots(k_j - 1)}$$
$$= \frac{k_1 \cdots k_j}{(k_1 - 1)\cdots(k_j - 1)} - 1 \,.$$

Hence, summing (3.5) and imposing $\sum_j \lambda_j = 1$ we find the equation

$$1 = \frac{1}{\mu - 1}\left(\frac{k_1 \cdots k_N}{(k_1 - 1)\cdots(k_N - 1)} - 1\right),$$

which determines uniquely μ as

$$\mu = \frac{k_1 \cdots k_N}{(k_1 - 1)\cdots(k_N - 1)} \,. \tag{3.6}$$

A second corollary of the computations in [27] is that

COROLLARY 3.4. — *Assume* $\{X_1, \ldots, X_N\} \in \mathbb{R}^{2\times 2}$ *induce the* T_N *configuration of form* (2.7) *and let* μ *and* λ *be as in* (3.5) *and* (3.6)*. Then* $A^\mu \lambda = 0$.

3.2. A characterization of T_N configurations in $\mathbb{R}^{n\times m}$

We start with a straightforward consequence of the results above.

Let us first introduce some notation concerning multi-indexes. We will use I for multi-indexes referring to ordered sets of rows of matrices and J for multi-indexes referring to ordered sets of columns. In our specific case, where we deal with matrices in $\mathbb{R}^{n\times m}$ we will thus have

$$I = (i_1, \ldots, i_r), \qquad 1 \leqslant i_1 < \cdots < i_r \leqslant n \,,$$
$$\text{and} \qquad J = (j_1, \ldots, j_s), \qquad 1 \leqslant j_1 < \cdots < j_s \leqslant m$$

and we will use the notation $|I| := r$ and $|J| := s$. In the sequel we will always have $r = s$.

DEFINITION 3.5. — *We denote by $\mathcal{A}_r$ the set*

$$\mathcal{A}_r = \{(I,J) : |I| = |J| = r\}, \qquad 2 \leqslant r \leqslant \min(n,m).$$

For a matrix $M \in \mathbb{R}^{n \times m}$ and for $Z \in \mathcal{A}_r$ of the form $Z = (I,J)$, we denote by M^Z the squared $r \times r$ matrix obtained by A considering just the elements a_{ij} with $i \in I$, $j \in J$ (using the order induced by I and J).

Given a set $\{X_1, \ldots, X_N\} \subset \mathbb{R}^{n \times m}$, $\mu \in \mathbb{R}$ and $Z \in \mathcal{A}_r$, we introduce the matrix

$$A_Z^\mu := \begin{bmatrix} 0 & \det(X_2^Z - X_1^Z) & \det(X_3^Z - X_1^Z) & \ldots & \det(X_N^Z - X_1^Z) \\ \mu \det(X_1^Z - X_2^Z) & 0 & \det(X_3^Z - X_2^Z) & \ldots & \det(X_N^Z - X_2^Z) \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \mu \det(X_1^Z - X_N^Z) & \mu \det(X_2^Z - X_N^Z) & \mu \det(X_3^Z - X_N^Z) & \ldots & 0 \end{bmatrix}.$$

PROPOSITION 3.6. — *A set $\{X_1, \ldots, X_N\} \subset \mathbb{R}^{n \times m}$ induces a T_N configuration if and only if there is a real $\mu > 1$ and a vector $\lambda \in \mathbb{R}^N$ with positive components such that*

$$A_Z^\mu \lambda = 0 \qquad \forall\, Z \in \mathcal{A}_2\,.$$

Moreover, if we define the vectors t^i as in (3.1), the coefficients k_i through (3.3) and the matrices P and C_i through (3.2), then $P, C_1, \ldots, C_N$ and $k_1, \ldots, k_N$ give a T_N configuration induced by $\{X_1, \ldots, X_N\}$.

For this reason and in view of Remark 3.3, we can introduce the following terminology:

DEFINITION 3.7. — *Given a T_N-configuration $P, C_1, \ldots, C_N$ and $k_1, \ldots, k_N$ we let μ and λ be given by (3.5) and (3.6) and we call $(\lambda, \mu) \in \mathbb{R}^{N+1}$ the defining vector of the T_N configuration.*

Proof of Proposition 3.6.

Direction $\Longleftarrow$. — Fix a set $\{X_1, \ldots, X_N\}$ of matrices with the property that there is a common $\mu > 1$ and a common λ with positive entries such that $A_Z^\mu \lambda = 0$ for every $Z \in \mathcal{A}_2$. For each Z we consider the corresponding set $\{X_1^Z, \ldots, Z_N^Z\}$ and we use the formulas (3.1), (3.3) and (3.2) to find $k_1, \ldots, k_N$, $P(Z)$ and $C_i(Z)$ such that

$$X_i^Z = P(Z) + C_1(Z) + \cdots + C_{i-1}(Z) + k_i C_i(Z)\,.$$

Since the coefficients k_i are independent of Z, the formulas give that the matrices $C_i(Z)$ (and $P(Z)$) are compactible, in the sense that, if $j\ell$ is an entry common to Z and Z', then $(C_i(Z))_{j\ell} = (C_i(Z'))_{j\ell}$. In particular there are matrics C_i's and P such that $C_i(Z) = C_i^Z$ and $P(Z) = P^Z$ and thus (2.7) holds. Moreover, we also know from Proposition 3.1 that $\text{rank}(C_i^Z) \leqslant 1$ for every Z and thus $\text{rank}(C_i) \leqslant 1$. We also know that $C_1^Z + \cdots + C_N^Z = 0$ for every Z and thus $C_1 + \cdots + C_N = 0$.

Direction $\implies$. — Assume $X_1, \ldots, X_N$ induce a T_N configuration as in (2.7). Then $X_1^Z, \ldots, X_N^Z$ induce a T_N configuration with corresponding $P^Z, C_1^Z, \ldots, C_N^Z$ and $k_1, \ldots, k_N$, where the latter coefficients are independent of Z. Thus, by Corollary 3.4, $A_Z^\mu \lambda = 0$. $\qquad\square$

3.3. Computing minors

We end this section with a further generalization, this time of (3.4): we want to extend the validity of it to any minor.

PROPOSITION 3.8. — *Let $\{X_1, \ldots, X_N\} \subset \mathbb{R}^{n \times m}$ induce a T_N configuration as in (2.7) with defining vector (λ, μ). Define the vectors $t^1, \ldots, t^N$ as in (3.1) and for every $Z \in \mathcal{A}_r$ of order $r \leqslant \min\{n, m\}$ define the minor $\mathcal{S} : \mathbb{R}^{n \times m} \ni X \mapsto \mathcal{S}(X) := \det(X^Z) \in \mathbb{R}$. Then*

$$\sum_{j=1}^{N} t_j^i \mathcal{S}(X_j) = \mathcal{S}\left(\sum_{j=1}^{N} t_j^i X_j\right) = \mathcal{S}(P + C_1 + \cdots + C_{i-1}). \qquad (3.7)$$

and $A_Z^\mu \lambda = 0$.

Fix any matrix $A \in \mathbb{R}^{m \times m}$. In the following we will denote by $\operatorname{cof}(A)$ the $m \times m$ matrix defined[3] as

$$\operatorname{cof}(A)_{ij} := (-1)^{i+j} \det(A^{j,i}),$$

where $A^{j,i}$ is the $m - 1 \times m - 1$ matrix obtained by eliminating from A the j-th row and the i-th column. It is well-known that

$$A \operatorname{cof}(A) = \operatorname{cof}(A)A = \det(A) \operatorname{Id}_m.$$

We will need the following elementary linear algebra fact, which in the literature is sometimes called Matrix Determinant Lemma:

LEMMA 3.9. — *Let A, B be matrices in $\mathbb{R}^{m \times m}$, and let $\operatorname{rank}(B) \leqslant 1$. Then,*

$$\det(A + B) = \det(A) + \langle \operatorname{cof}(A)^T, B \rangle$$

Moreover, we need another elementary computation, which is essentially contained in [27] and for which we report the proof at the end of the section for the reader's convenience.

---

[3] Note that sometimes in the literature one refers to what we called $\operatorname{cof}(A)$ as the *adjoint* of A, and the adjoint of A would be $\operatorname{cof}(A)^T$.

LEMMA 3.10. — *Assume the real numbers $\mu > 1$, $\lambda_1, \ldots, \lambda_N > 0$ and $k_1, \ldots, k_N > 1$ are linked by the formulas (3.3). Assume $v, v_1, \ldots, v_N$, $w_1, \ldots, w_N$ are elements of a vector space satisfying the relations*

$$w_i = v + v_1 + \cdots + v_{i-1} + k_i v_i \qquad (3.8)$$

$$0 = v_1 + \cdots + v_N. \qquad (3.9)$$

If we define the vectors t^i as in (3.1), then

$$\sum_j t^i_j w_j = v + v_1 + \cdots + v_{i-1}. \qquad (3.10)$$

Proof of Proposition 3.8. — Fix the Z of the statement of the proposition. $X_1^Z, \ldots, X_N^Z$ induces T_N with the same coefficients $k_1, \ldots k_N$. This reduces therefore the statement to the case in which $m = n$, $Z = ((1, \ldots n), (1, \ldots, n))$ and the minor $\mathcal{S}$ is the usual determinant.

We first prove (3.7). In order to do this we specialize (3.10) to $w_\ell = \det(X_\ell)$, $v = \det(P)$, $v_\ell = \langle \operatorname{cof}^T(P + C_1 + \cdots + C_{\ell-1}), C_\ell \rangle$. To simplify the notation set

$$P^{(1)} = P, \quad \text{and} \quad P^{(\ell)} = P + C_1 + \cdots + C_{\ell-1} \qquad \forall \, \ell \in \{1, \ldots, N+1\}.$$

We want to show that

$$v + v_1 + \cdots + v_{i-1} = \det(P^{(i)}) \quad \text{and} \quad v_1 + \cdots + v_N = 0,$$

and this would conclude the proof of (3.7) because of Lemma 3.10. A repeated application of Lemma 3.9 yields:

$$v + v_1 + \cdots + v_{i-1}$$
$$= \underbrace{\underbrace{\det(P) + \langle \operatorname{cof}^T(P), C_1 \rangle}_{\det(P^{(2)})} + \langle \operatorname{cof}^T(P^{(2)}), C_2 \rangle + \cdots + \langle \operatorname{cof}^T(P^{(i)}), C_{i-1} \rangle}_{\det(P^{(3)})}$$
$$= \det(P^{(i)}) = \det(P + C_1 + \cdots + C_{i-1}).$$

As a consequence of Lemma 3.9, we also have $v_\ell = \det(P^{(\ell+1)}) - \det(P^{(\ell)})$. Therefore:

$$v_1 + \cdots + v_N = \sum_{\ell=1}^{N} \left(\det(P^{(\ell+1)}) - \det(P^{(\ell)}) \right) = \det(P^{(N+1)}) - \det(P^{(1)}).$$

Since $\sum_\ell C_\ell = 0$ and $\det(P^{(N+1)}) = \det(P + \sum_\ell C_\ell)$, the following holds

$$\det(P^{(N+1)}) - \det(P^{(1)}) = \det\left(P + \sum_\ell C_\ell \right) - \det(P) = \det(P) - \det(P) = 0,$$

and the conclusion is thus reached.

To prove the second part of the statement notice that $A^\mu_Z \lambda = 0$ is equivalent to the following N equations:

$$\sum_{j=1}^{N} t^i_j \det(X_j - X_i) = 0 \qquad \forall\, i \in \{1, \ldots, N\}.$$

Fix $i \in \{1, \ldots, N\}$ and define matrices $Y_j := X_j - X_i,\, \forall\, j.$ $\{Y_1, \ldots, Y_N\}$ is still a T_N configuration of the form

$$Y_i = P' + \sum_{\ell=1}^{i-1} C_\ell + k_i C_i,$$

and $P' = -X_i$ (recall that $P = 0$). Apply now (3.7) to find that

$$\sum_j t^i_j \det(X_j - X_i) = \sum_j t^i_j \det(Y_j)$$

$$= \det\left(P' + \sum_{\ell=1}^{i-1} C_\ell\right) = \det\left(-X_i + \sum_{\ell=1}^{i-1} C_\ell\right) = \det(-k_i C_i) = 0$$

and conclude the proof. □

3.4. Proof of Lemma 3.10

It is sufficient to compute separately $\sum_{j=1}^{N} t^1_j w_j = \sum_{j=1}^{N} \lambda_j w_j$ and $\sum_{j=1}^{i-1} \lambda_j w_j$. In fact,

$$\sum_j^{N} t^i_j w_j = \frac{1}{\xi_i}\left[\sum_{j=1}^{N} \lambda_j w_j + (\mu - 1)\sum_{j=1}^{i-1} \lambda_j w_j\right]. \tag{3.11}$$

We can write

$$\sum_j \lambda_j w_j = v + a_1 v_1 + \cdots + a_N v_N,$$

being, $\forall\, \ell \in \{1, \ldots, N\}, a_\ell = k_\ell \lambda_\ell + \cdots + \lambda_N$. Recalling that the defining vector and the numbers k_i are related through (3.3), we compute

$$a_\ell = k_\ell \lambda_\ell + \cdots + \lambda_N = \frac{\mu \lambda_1 + \cdots + \mu \lambda_\ell + \lambda_{\ell+1} + \ldots \lambda_N}{\mu - 1} + \lambda_{\ell+1} + \cdots + \lambda_N$$

$$= \frac{\mu(\lambda_1 + \cdots + \lambda_N)}{\mu - 1} = \frac{\mu}{\mu - 1} =: a. \tag{3.12}$$

Hence

$$\sum_{j=1}^{N} \lambda_j w_j = v + \frac{\mu}{\mu - 1}(v_1 + \cdots + v_N).$$

On the other hand,

$$\sum_{j=1}^{i-1} \lambda_j w_j = b_1 v + b_2 v_1 + \cdots + b_i v_{i-1},$$

and

$$b_1 = \lambda_1 + \cdots + \lambda_{i-1} =: c,$$

$$b_\ell = k_\ell \lambda_\ell + \cdots + \lambda_{i-1} = \frac{\mu(\lambda_1 + \cdots + \lambda_\ell) + \sum_{j=\ell+1}^{N} \lambda_j + (\mu - 1)\sum_{j=\ell+1}^{i-1} \lambda_j}{\mu - 1}$$

$$= \frac{\mu(\sum_{j=1}^{i-1} \lambda_j) + \sum_{j=i}^{N} \lambda_j}{\mu - 1} =: b, \quad \forall \ell \in \{2, \ldots, i\}.$$

Also,

$$\xi_i = \|(\mu\lambda_1, \ldots, \mu\lambda_{i-1}, \lambda_i, \ldots, \lambda_N)\|_1 = (\mu - 1)(\lambda_1 + \cdots + \lambda_{i-1}) + 1$$
$$= (\mu - 1)b = 1 + (\mu - 1)c.$$

We can now compute (3.11):

$$\frac{1}{\xi_i}\left[\sum_{j=1}^{N} \lambda_j w_j + (\mu - 1)\sum_{j=1}^{i-1} \lambda_j w_j\right]$$

$$= \frac{1}{\xi_i}\left[v + a_1 v_1 + \cdots + a_N v_N + (\mu - 1)(b_1 v + b_2 v_1 + \cdots + b_i v_{i-1})\right]$$

$$= \frac{1}{\xi_i}\left[(\mu - 1)b(v + v_1 + \cdots + v_{i-1}) + a(v_1 + \cdots + v_N)\right]$$

$$= v + v_1 + \cdots + v_{i-1} + \frac{a}{(\mu - 1)b}(v_1 + \cdots + v_N)$$

We use the just obtained identity

$$\sum_{j=1}^{N} t_j^i w_j = v + v_1 + \cdots + v_{i-1} + \frac{a}{(\mu - 1)b}(v_1 + \cdots + v_N) \qquad (3.13)$$

Using that $v_1 + \cdots + v_N = 0$ we conclude the desired identity.

4. Inclusions sets relative to polyconvex functions

In this section we consider the following question. Given a set of distinct matrices $A_i \in \mathbb{R}^{2n} \times \mathbb{R}^m$

$$A_i := \begin{pmatrix} X_i \\ Y_i \end{pmatrix}, \qquad (4.1)$$

do they belong to a set of the form

$$K'_f := \begin{pmatrix} X_i \\ Df(X_i) \end{pmatrix} \tag{4.2}$$

for some strictly polyconvex function $f : \mathbb{R}^{n \times m} \to \mathbb{R}$? In order to answer, we first need to introduce the following notation. Let $f : \mathbb{R}^{n \times m} \to \mathbb{R}$ be a strictly polyconvex function of the form $f(X) = g(\Phi(X))$, where $g \in C^1(\mathbb{R}^k)$ is strictly convex and Φ is the vector of all the subdeterminants of X, i.e.

$$\Phi(X) = (X, v_1(X), \ldots, v_{\min(n,m)}(X)),$$

and

$$v_s(X) = (\det(X_{Z_1}), \ldots, \det(X_{Z_{\#\mathcal{A}_s}}))$$

for some fixed (but arbitrary) ordering of all the elements $Z \in \mathcal{A}_s$. Variables of $\mathbb{R}^k$, and hence partial derivatives in $\mathbb{R}^k$, are labeled using the ordering induced by Φ. The first nm partial derivatives, corresponding in $\Phi(X)$ to X, are collected in a $n \times m$ matrix denoted with $D_X g$. The j-th partial derivative, $mn + 1 \leqslant j \leqslant k$, is instead denoted by $\partial_Z g$, where Z is the element of $\mathcal{A}_s$ corresponding to the j-th position of Φ. Let us make an example in low dimension: if $n = 3, m = 2$, then $k = 9$, and we choose the ordering of Φ to be

$$\Phi(X) = (X, \det(X_{(12,12)}), \det(X_{(13,12)}), \det(X_{(23,12)})).$$

In this case, $y \in \mathbb{R}^k$ has coordinates

$$y = (y_{11}, y_{12}, y_{21}, y_{22}, y_{31}, y_{32}, y_{(12,12)}, y_{(13,12)}, y_{(23,12)}).$$

The partial derivatives with respect to the first 6 variables are collected in the 3×2 matrix:

$$D_X g = \begin{pmatrix} \partial_{11} g & \partial_{12} g \\ \partial_{21} g & \partial_{22} g \\ \partial_{31} g & \partial_{32} g \end{pmatrix}$$

The partial derivatives with respect to the remaining variables are denoted as $\partial_{(12,12)} g$, $\partial_{(13,12)} g$ and $\partial_{(23,12)} g$, i.e. following the ordering induced by Φ. We are ready to state the following

PROPOSITION 4.1. — *If $A_i \in K'_f$ and $A_i \neq A_j$ for $i \neq j$, then X_i, $Y_i = Df(X_i)$ and $c_i = f(X_i)$ fulfill the following inequalities for every $i \neq j$:*

$$c_i - c_j + \langle Y_i, X_j - X_i \rangle$$

$$- \sum_{r=2}^{\min(m,n)} \sum_{Z \in \mathcal{A}_r} d_Z^i \big(\langle \operatorname{cof}(X_i^Z)^T, X_j^Z - X_i^Z \rangle - \det(X_j^Z) + \det(X_i^Z) \big) < 0, \tag{4.3}$$

where $d_Z^i = \partial_Z g(\Phi(X_i))$.

The expressions in (4.3) can be considerably simplied when the matrices $X_1, \ldots, X_N$ induce a T_N configuration.

LEMMA 4.2. — *Assume $X_1, \ldots, X_N$ induces a T_N configuration of the form* (2.7) *and associated vectors t^i, $i \in \{1, \ldots, N\}$. Then, $\forall\, i \in \{1, \ldots, N\}$, $\forall\, r \in \{2, \ldots, \min(m, n)\}$, $\forall\, Z \in \mathcal{A}_r$,*

$$\sum_j t^i_j \left(\langle \mathrm{cof}(X^Z_i)^T, X^Z_j - X^Z_i \rangle - \det(X^Z_j) + \det(X^Z_i) \right) = 0. \qquad (4.4)$$

In particular combining (4.3) and (4.4) we immediately get the following:

COROLLARY 4.3. — *Let f be a strictly polyconvex function and let $A_1, \ldots, A_N$ be distinct elements of K'_f with the additional property that $\{X_1, \ldots, X_N\}$ induces a T_N configuration of the form* (2.7) *with defining vector (μ, λ). Then,*

$$c_i - \sum_j t^i_j c_j - k_i \langle Y_i, C_i \rangle < 0, \qquad (4.5)$$

where the t^i's are given by (3.1).

4.1. Proof of Proposition 4.1

Since $A_i \neq A_j$, for $i \neq j$, we also have $X_i \neq X_j$, for $i \neq j$. Therefore, the strict convexity of g yields, for $i \neq j$,

$$\langle Dg(\Phi(X_i)), \Phi(X_j) - \Phi(X_i) \rangle < g(\Phi(X_j)) - g(\Phi(X_i)). \qquad (4.6)$$

A simple computation shows that for the function $\det(\,\cdot\,) : \mathbb{R}^{r \times r} \to \mathbb{R}$:

$$D(\det(X))|_{X=Y} = \mathrm{cof}(Y)^T.$$

In the following equation, we will write, for an $n \times m$ matrix M and for $Z \in \mathcal{A}_r$, $\overline{\mathrm{cof}(M^Z)^T}$ to denote the $n \times m$ matrix with 0 in every entry, except for the rows and columns corresponding to the multiindex $Z = (I, J)$, which will be filled with the entries of the matrix $\mathrm{cof}(M^Z)^T \in \mathbb{R}^{r \times r}$, namely, if $i \notin I$ or $j \notin J$, then $(\overline{\mathrm{cof}(M^Z)^T})_{ij} = 0$ and, if we eliminate all such coefficients, the remaining $r \times r$ matrix equals $\mathrm{cof}(M^Z)^T$. Using the notation on the derivatives of the function g introduced at the beginning of Section 4, we can write the following formula for the derivatives of f

$$Df(X) = D(g(\Phi(X))) = D_X g(\Phi(X)) + \sum_{r=2}^{\min(m,n)} \sum_{Z \in \mathcal{A}_r} \partial_Z g(\Phi(X)) \overline{\mathrm{cof}(X^Z)^T}.$$

When evaluated on $X = X_i$,

$$Y_i = D_X g(\Phi(X_i)) + \sum_{r=2}^{\min(m,n)} \sum_{Z \in \mathcal{A}_r} \partial_Z g(\Phi(X_i)) \overline{\mathrm{cof}(X_i^Z)}$$

In order to simplify the notation set now $d_Z^i := \partial_Z g(\Phi(X_i))$. The previous expression yields:

$$\langle Dg(\Phi(X_i)), \Phi(X_j) - \Phi(X_i) \rangle$$

$$= \langle D_X g(\Phi(X_i)), X_j - X_i \rangle + \sum_{r=2}^{\min(m,n)} \sum_{Z \in \mathcal{A}_r} d_Z^i \left(\det(X_j^Z) - \det(X_i^Z) \right)$$

$$= \left\langle Y_i - \sum_{r=2}^{\min(m,n)} \sum_{Z \in \mathcal{A}_r} d_Z^i \overline{\mathrm{cof}(X_i^Z)^T}, X_j - X_i \right\rangle$$

$$+ \sum_{r=2}^{\min(m,n)} \sum_{Z \in \mathcal{A}_r} d_Z^i \left(\det(X_j^Z) - \det(X_i^Z) \right).$$

Since

$$g(\Phi(X_j)) - g(\Phi(X_i)) = f(X_j) - f(X_i) = c_j - c_i,$$

(4.6) becomes:

$$\langle Y_i, X_j - X_i \rangle$$
$$- \sum_{r=2}^{\min(m,n)} \sum_{Z \in \mathcal{A}_r} d_Z^i \left(\langle \overline{\mathrm{cof}(X_i^Z)^T}, X_j - X_i \rangle - \det(X_j^Z) + \det(X_i^Z) \right) < c_j - c_i.$$

Finally, summing $c_i - c_j$ on both sides:

$$c_i - c_j + \langle Y_i, X_j - X_i \rangle$$
$$- \sum_{r=2}^{\min(m,n)} \sum_{Z \in \mathcal{A}_r} d_Z^i \left(\langle \overline{\mathrm{cof}(X_i^Z)^T}, X_j - X_i \rangle - \det(X_j^Z) - \det(X_i^Z) \right) < 0 \quad (4.7)$$

Using the fact that $\langle \overline{\mathrm{cof}(X_i^Z)^T}, X_j - X_i \rangle = \langle \mathrm{cof}(X_i^Z)^T, X_j^Z - X_i^Z \rangle$, we see that the previous inequality implies the conclusion

$\forall\, i \neq j$,

$$c_i - c_j + \langle Y_i, X_j - X_i \rangle$$
$$- \sum_{r=2}^{\min(m,n)} \sum_{Z \in \mathcal{A}_r} d_Z^i \left(\langle \mathrm{cof}(X_i^Z)^T, X_j^Z - X_i^Z \rangle - \det(X_j^Z) + \det(X_i^Z) \right) < 0.$$

4.2. Proof of Lemma 4.2

The result is a direct consequence of Lemma 3.9 and Proposition 3.8. First of all, by Proposition 3.8 we have

$$\sum_j t^i_j \det(X^Z_j) = \det\left(\sum_j t^i_j X^Z_j\right) = \det\left(P^Z_1 + \cdots + C^Z_{i-1}\right) \qquad (4.8)$$

Moreover, by (3.2), we get

$$\sum_j t^i_j \langle \operatorname{cof}(X^Z_i)^T, X^Z_j - X^Z_i \rangle$$
$$= \langle \operatorname{cof}(X^Z_i)^T, P^Z + C^Z_1 + \cdots + C^Z_{i-1} - X^Z_i \rangle \qquad (4.9)$$
$$= -k_i \langle \operatorname{cof}(X^Z_i)^T, C^Z_i \rangle.$$

Finally, apply Lemma 3.9 to $A = X^Z_i$ and $B = -k_i C^Z_i$ to get

$$\det(P^Z + \cdots + C^Z_{i-1}) = \det(X^Z_i) - k_i \langle \operatorname{cof}(X^Z_i)^T, C^Z_i \rangle. \qquad (4.10)$$

These three equalities together give (4.4).

4.3. Proof of Corollary 4.3

Multiply (4.3) by t^i_j and sum over j. Using Lemma 4.2 and taking into account $\sum_j t^i_j = 1$ we get

$$c_i - \sum_j t^i_j c_j + \left\langle Y_i, \sum_j t^i_j X_j - X_i \right\rangle < 0.$$

Since

$$\sum_j t^i_j X_j = P + C_1 + \cdots + C_{i-1}$$

and

$$X_i = P + C_1 + \cdots + C_{i-1} + k_i C_i,$$

we easily conclude (4.5).

5. Proof of Theorem 1.2

In this section we prove the main theorem of this paper.

5.1. Gauge invariance

In the first part we state a corollary of some obvious invariance of poly-convex functions under certain groups of transformations. Indeed, despite the technical details of the proof, this Lemma stems from the simple observations that if $f : \mathbb{R}^{n \times m} \to \mathbb{R}$ is strictly polyconvex, then

$$\bar{f}(X) = f(X + M) + \langle X, N \rangle + c$$

is still strictly polyconvex, for any $M, N \in \mathbb{R}^{n \times m}$ and $c \in \mathbb{R}$, and that T'_N configurations are invariant under translation. This invariance will then be used in the proof of Theorem 1.2 to bring an hypothetical T'_N configuration into a "canonical form".

LEMMA 5.1. — *Let $f : \mathbb{R}^{n \times m}$ be strictly polyconvex and assume that K_f contains a set of matrices $\{A_1, \ldots, A_N\}$ which induces a nondegenerate T'_N configuration, denoted by*

$$A_i := \begin{pmatrix} \overline{X}_i \\ \overline{Y}_i \\ \overline{Z}_i \end{pmatrix},$$

where

$$\overline{X}_i = P + \overline{C}_1 + \cdots + k_i \overline{C}_i,$$
$$\overline{Y}_i = Q + \overline{D}_1 + \cdots + k_i \overline{D}_i,$$
$$\overline{Z}_i = R + \overline{E}_1 + \cdots + k_i \overline{E}_i.$$

Then, for every $S, T \in \mathbb{R}^{n \times m}, a \in \mathbb{R}$, there exists another strictly polyconvex function $\bar{f}$ such that the family of matrices

$$B_i := \begin{pmatrix} X_i \\ Y_i \\ Z_i \end{pmatrix}$$

lie in $K_{\bar{f}}, \forall\, i$, and they have the following properties:

- *The matrices X_i, Y_i have the form*
$$X_i = S + \overline{C}_1 + \cdots + k_i \overline{C}_i,$$
$$Y_i = T + \overline{D}_1 + \cdots + k_i \overline{D}_i.$$

- *the matrices Z_i are of the form*
$$Z_i = U + E_1 + \cdots + k_i E_i,$$

  where

$$U = R - P^T(Q - T) + (S - P)^T NT + (\langle P, Q - T \rangle + a) \,\mathrm{id}\,.$$

Moreover, if $n_i \in \mathbb{R}^m$ is the unit vector of Proposition 2.7, then we have

$$\sum_i E_i = 0, \tag{5.1}$$

$$\sum_j t_j^i Z_i = U + E_1 + \cdots + E_{i-1}, \quad \forall\, i \in \{1, \ldots, N\}, \tag{5.2}$$

$$E_i n_i = 0, \quad \forall\, i. \tag{5.3}$$

Proof. — We consider $\bar{f}$ of the form

$$f(X) = \bar{f}(X + O) + \langle X, V \rangle + a.$$

We want $X_i - \overline{X}_i = S - P$ and $Y_i - \overline{Y}_i = T - Q$, therefore the natural choice for O is $O := -P + S$. In this way,

$$f(\overline{X}_i) = \bar{f}(\overline{X}_i - P + S) + \langle \overline{X}_i, V \rangle + a = \bar{f}(X_i) + \langle \overline{X}_i, V \rangle + a.$$

Moreover, we have

$$Df(\overline{X}_i) = \overline{Y}_i,$$

hence $D\bar{f}(X_i) = Y_i$ if and only if

$$Df(\overline{X}_i) - V = \overline{Y}_i - V = Y_i,$$

i.e. $V := Q - T$. We now show that the modification of $\overline{Z}_i$ into Z_i with the properties listed in the statement of the present proposition will let us fulfill also the last requirement, namely that

$$Z_i = X_i^T Y_i - c_i'\, \mathrm{id},$$

where $c_i' = \bar{f}(X_i)$. Analogously, we denote with $c_i = f(\overline{X}_i)$. We write

$$\begin{aligned}
X_i^T Y_i - c_i'\, \mathrm{id} &= (X_i - O)^T Y_i + O^T Y_i - (c_i - \langle \overline{X}_i, V \rangle - a)\, \mathrm{id} \\
&= \overline{X}_i^T (Y_i + V) - \overline{X}_i^T V + O^T Y_i - (c_i - \langle \overline{X}_i, V \rangle - a)\, \mathrm{id} \\
&= \underbrace{\overline{X}_i^T \overline{Y}_i - c_i\, \mathrm{id}}_{\tilde{Z}_i} - \overline{X}_i^T V + O^T Y_i + (\langle \overline{X}_i, V \rangle + a)\, \mathrm{id}.
\end{aligned}$$

We can thus rewrite

$$\overline{X}_i^T V = P^T V + \sum_{j=1}^{i-1} C_j^T V + k_i C_i^T V.$$

For every fixed j, we decompose in a unique way $V = V_j + V_j^\perp$, where $V_j^\perp = (V n_j) \otimes n_j$ and $V_j = V - V_j^\perp$. Note that, since $C_j = u_j \otimes v_j$, this

implies that $C_j^T V_j - \langle C_j, V_j \rangle$ id is a scalar multiple of the orthogonal projection on $\operatorname{span}(n_j)^\perp$. Therefore,

$$C_j^T V = C_j^T V_j + C_j^T V_j^\perp = \underbrace{C_j^T V_j - \langle C_j, V_j \rangle \,\mathrm{id} + C_j^T V_j^\perp}_{=:R_j} + \langle C_j, V_j \rangle \,\mathrm{id}\,.$$

Consequently, $\overline{X}_i^T V$ has the following form:

$$\overline{X}_i^T V = P^T V + \sum_{j=1}^{i-1} R_j + k_i R_i + \left(\sum_{j=1}^{i-1} \langle C_j, V_j \rangle + k_i \langle C_i, V_i \rangle \right) \mathrm{id}\,.$$

Finally, we define, $Z_i' := P^T V + \sum_{j=1}^{i-1} R_j + k_i R_i$. Resuming the main computations, we have obtained that:

$$X_i^T Y_i - c_i' \,\mathrm{id}$$

$$= \overline{Z}_i - Z_i' + O^T Y_i + \left(-\sum_{j=1}^{i-1} \langle C_j, V_j \rangle - k_i \langle C_i, V_i \rangle + \langle \overline{X}_i, V \rangle + a \right) \mathrm{id}\,.$$

Since

$$-\sum_{j=1}^{i-1} \langle C_j, V_j \rangle - k_i \langle C_i, V_i \rangle + \langle \overline{X}_i, V \rangle$$

$$= -\sum_{j=1}^{i-1} \langle C_j, V \rangle - k_i \langle C_i, V \rangle + \langle \overline{X}_i, V \rangle = \langle P, V \rangle,$$

we are finally able to say that the first part of the Proposition is proved provided that

$$Z_i := \overline{Z}_i - Z_i' + O^T Y_i + (\langle P, V \rangle + a) \,\mathrm{id},$$
$$E_i := \overline{E}_i - C_i^T V_i + \langle C_i, V_i \rangle \,\mathrm{id} - C_i^T V_i^\perp + O^T D_i,$$
$$U := R - P^T V + O^T T + (\langle P, V \rangle + a) \,\mathrm{id}\,.$$

To simplify future computations, let us use the identities $V_i + V_i^\perp = V$ and $\langle C_i, V_i \rangle = \langle C_i, V \rangle$:

$$Z_i := \overline{Z}_i - Z_i' + O^T Y_i + (\langle P, V \rangle + a) \,\mathrm{id},$$
$$E_i := \overline{E}_i - C_i^T V + \langle C_i, V \rangle \,\mathrm{id} + O^T D_i,$$
$$U := R - P^T V + O^T T + (\langle P, V \rangle + a) \,\mathrm{id}\,.$$

Properties (5.1), (5.2) and (5.3) are easily checked by the linearity of the previous expressions and the identity (3.13). □

5.2. Proof of Theorem 1.2

Assume by contradiction the existence of a T'_N configuration induced by matrices $\{A_1, \ldots, A_N\}$ which belong to the inclusion set K_f of some stictly polyconvex function $f \in C^1(\mathbb{R}^{n \times m})$. Note that the corresponding $\{X_i\}$ must be all distinct, because $Y_i = Df(X_i)$ and $Z_i = X_i^T Df(X_i) - f(X_i) \,\mathrm{id}$. Thus $\{X_1, \ldots, X_N\}$ induce a T_N configuration.

We consider coefficients $k_1, \ldots, k_N$ and matrices P, Q, R, C_i, D_i, E_i as in Definition 2.6. By Lemma 5.1 we can assume, without loss of generality, that

$$P = 0 = Q \quad \text{and} \quad \mathrm{tr}(R) = 0 \,.$$

We are now going to prove that the system of inequalities

$$- \nu_i := c_i - \sum_j t_j^i c_j - k_i \langle Y_i, C_i \rangle < 0, \quad \forall\, i, \tag{5.4}$$

where c_i and t_j^i are as in Corollary 4.3, cannot be fulfilled at the same time. This will then give a contradiction.

A rough idea of the proof is the following. As explained in Subsection 2.4, in order to reach a contradiction we need to use Z_i in an essential way. In particular, we need to exploit the fact that

$$Z_i = X_i^T Y_i - c_i \,\mathrm{id}_m$$

and that

$$Z_i = R + E_1 + \cdots + k_i E_i.$$

Since the only information we have on E_i is that $E_i n_i = 0$, the first thing we do is to use the structure of Z_i to rewrite E_i in terms of $C_1, \ldots, C_N$, $D_1, \ldots, D_N, c_1, \ldots, c_N$, and then exploit the information $E_i n_i = 0$ to gain additional conditions on $C_1, \ldots, C_N, D_1, \ldots, D_N, c_1, \ldots, c_N$. This is done in (5.5)–(5.11). This leads to the fact that $\xi_i \nu_i, \xi > 0$, are "generalized eigenvalues" of the trace-free matrix R, compare (5.12). This gives the required contradiction.

In order to follow our strategy, we need to compute the following sums:

$$\sum_j t_j^i Z_j = \sum_j t_j^i X_j^T Y_j - \sum_j t_j^i c_j \,\mathrm{id} \,. \tag{5.5}$$

Let us start computing the sum for $i = 1$, $\sum_j \lambda_j X_j^T Y_j$. We rewrite it in the following way:

$$\sum_j \lambda_j X_j^T Y_j$$

$$= \sum_{j=1}^{N} \lambda_j \left(\sum_{1 \leqslant a,b \leqslant j-1} C_a^T D_b + k_j \sum_{1 \leqslant a \leqslant j-1} C_a^T D_j + k_j \sum_{1 \leqslant b \leqslant j-1} C_j^T D_b + k_j^2 C_j^T D_j \right)$$

$$= \sum_{i,j} g_{ij} C_i^T D_j, \tag{5.6}$$

where we collected in the coefficients g_{ij} the following quantities:

$$g_{ij} = \begin{cases} \lambda_i k_i + \sum_{r=i+1}^{N} \lambda_r, & \text{if } i \neq j \\ \lambda_i k_i^2 + \sum_{r=i+1}^{N} \lambda_r, & \text{if } i = j. \end{cases}$$

As already computed in (3.12), we have:

$$g_{ij} = g_{ji} = \lambda_i k_i + \sum_{r=i+1}^{N} \lambda_r = \frac{\mu}{\mu - 1},$$

On the other hand,

$$g_{ii} = k_i^2 \lambda_i + \sum_{r=i+1}^{N} \lambda_r = k_i(k_i - 1)\lambda_i + \frac{\mu}{\mu - 1}.$$

Using the equalities $\sum_\ell C_\ell = 0 = \sum_\ell D_\ell$, then also $\sum_{i,j} C_i^T D_j = 0$, and so $\sum_{i \neq j} C_i^T D_j = -\sum_i C_i^T D_i$. Hence, (5.6) becomes

$$\sum_{i,j} g_{ij} C_i^T D_j = \frac{\mu}{\mu - 1} \sum_{i \neq j} C_i^T D_j + \sum_i \left(k_i(k_i - 1)\lambda_i + \frac{\mu}{\mu - 1} \right) C_i^T D_i$$

$$= \sum_i k_i(k_i - 1)\lambda_i C_i^T D_i.$$

We just proved that

$$\sum_j \lambda_j X_j^T Y_j = \sum_i k_i(k_i - 1)\lambda_i C_i^T D_i. \tag{5.7}$$

In particular,

$$\sum_j \lambda_j \langle X_j, Y_j \rangle = 0, \tag{5.8}$$

since $C_i^T D_i$ is trace-free for every i. We also have:

$$\sum_j \lambda_j Z_j = \sum_j \lambda_j X_j^T Y_j - \sum_j \lambda_j c_j \, \mathrm{id}$$

$$\implies \sum_i k_i(k_i - 1)\lambda_i C_i^T D_i = R + \sum_j \lambda_j c_j \, \mathrm{id}.$$

Since both $\mathrm{tr}(R)$ and $\mathrm{tr}\left(\sum_i k_i(k_i - 1)\lambda_i C_i^T D_i\right) = 0$, then $\sum_j \lambda_j c_j = 0$ and we get

$$\sum_i k_i(k_i - 1)\lambda_i C_i^T D_i = R.$$

Recall the definition of t^i, namely

$$t^i = \frac{1}{\xi_i}\left(\mu\lambda_1, \ldots, \mu\lambda_{i-1}, \lambda_i, \ldots, \lambda_N\right).$$

By the previous computation $(i = 1)$ and (5.2), it is convenient to rewrite (5.5) as

$$R + \sum_j^{i-1} E_j = \frac{1}{\xi_i}\left(R + (\mu - 1)\sum_{j=1}^{i-1}\lambda_j X_j^T Y_j\right) - \sum_j t_j^i c_j \, \mathrm{id}. \qquad (5.9)$$

Once again, let us express the sum up to $i - 1$ in the following way:

$$\sum_{j=1}^{i-1}\lambda_j X_j^T Y_j = \sum_{k,j}^{i-1} s_{kj} C_k^T D_j.$$

A combinatorial argument analogous to the one in the previous case gives

$$s_{\ell\ell} = k_\ell^2 \lambda_\ell + \cdots + \lambda_{i-1} = (k_\ell^2 - k_\ell)\lambda_\ell + k_\ell\lambda_\ell + \cdots + \lambda_{i-1},$$
$$s_{\alpha\beta} = k_\alpha\lambda_\alpha + \cdots + \lambda_{i-1}, \qquad \alpha > \beta$$
$$s_{\beta\alpha} = k_\beta\lambda_\beta + \cdots + \lambda_{i-1}, \qquad \alpha < \beta.$$

Now

$$k_r\lambda_r + \cdots + \lambda_{i-1} = \frac{\mu(\sum_{j=1}^{i-1}\lambda_j) + \sum_{j=i}^{N}\lambda_j}{\mu - 1}$$

and so

$$k_r\lambda_r + \cdots + \lambda_{i-1} = \frac{(\mu - 1)(\sum_{j=1}^{i-1}\lambda_j) + 1}{\mu - 1} = \frac{\xi_i}{\mu - 1} =: b_{i-1}$$

Hence

$$\sum_{j=1}^{i-1}\lambda_j X_j^T Y_j = \sum_{k,j}^{i-1} s_{kj} C_k^T D_j = b_{i-1}\sum_{k,j}^{i-1} C_k^T D_j + \sum_{\alpha=1}^{i-1} k_\alpha(k_\alpha - 1)\lambda_\alpha C_\alpha^T D_\alpha$$

We rewrite (5.9) as

$$R + \sum_{j=1}^{i-1} E_j = \frac{1}{\xi_i} \left(R + \xi_i \sum_{k,j}^{i-1} C_k^T D_j + (\mu - 1) \sum_{\alpha=1}^{i-1} k_\alpha (k_\alpha - 1) \lambda_\alpha C_\alpha^T D_\alpha \right)$$
$$- \sum_j t_j^i c_j \, \mathrm{id} \quad (5.10)$$

E_i is readily computed using (5.10) and the definition of Z_i:

$$k_i E_i + \frac{1}{\xi_i} \left(R + \xi_i \sum_{k,j}^{i-1} C_k^T D_j + (\mu - 1) \sum_{\alpha=1}^{i-1} k_\alpha (k_\alpha - 1) \lambda_\alpha C_\alpha^T D_\alpha \right) - \sum_j t_j^i c_j \, \mathrm{id}$$
$$= X_i^T Y_i - c_i \, \mathrm{id}$$

then

$$k_i E_i + \frac{1}{\xi_i} \left(R + (\mu - 1) \sum_{\alpha=1}^{i-1} k_\alpha (k_\alpha - 1) \lambda_\alpha C_\alpha^T D_\alpha \right)$$
$$= k_i \sum_j^{i-1} C_i^T D_j + k_i \sum_j^{i-1} C_j^T D_i + k_i^2 C_i^T D_i - c_i \, \mathrm{id} + \sum_j t_j^i c_j \, \mathrm{id} \,.$$

The evaluation of the previous expression at the vectors n_i of Proposition 2.7(ii) yields

$$\frac{1}{\xi_i} \left(R n_i + (\mu - 1) \sum_{\alpha=1}^{i-1} k_\alpha (k_\alpha - 1) \lambda_\alpha C_\alpha^T D_\alpha n_i \right)$$
$$= k_i \sum_j^{i-1} C_i^T D_j n_i - c_i n_i + \sum_j t_j^i c_j n_i. \quad (5.11)$$

Now, since $C_i v = 0$, $\forall \, v \perp n_i$, we must have

$$C_i^T D_j n_i = \langle C_i, D_j \rangle n_i.$$

The last equality implies that the right hand side of (5.11) is exactly $\nu_i n_i$, where ν_i has been defined in (5.4). We will now prove that there exists a nontrivial subset $A \subset \{1, \ldots, N\}$ such that

$$\sum_{j \in A} \xi_j \nu_j = 0,$$

and this will conclude the proof, being $\xi_j > 0$, $\forall \, j$. Since $C_\alpha^T D_\alpha n_i = b_{\alpha i} n_\alpha$, if we define

$$a_{\alpha i} := -(\mu - 1) \xi_i k_\alpha (k_\alpha - 1) \lambda_\alpha b_{\alpha i} \,,$$

then we can rewrite (5.11) as

$$Rn_i = \xi_i \nu_i n_i + \sum_{\alpha=1}^{i-1} a_{\alpha i} n_\alpha.$$

Now, consider the set $A \subset \{1, \ldots, N\}$ defined as

$$A = \{1\} \cup \left\{ j : \begin{array}{l} n_j \text{ cannot be written as a linear combination} \\ \text{of vectors } n_\ell, \text{ for any } \ell \leqslant j \end{array} \right\}.$$

Clearly

$$\mathrm{span}(\{n_s : s \in A\}) = \mathrm{span}(n_1, \ldots, n_N) \subset \mathbb{R}^m$$

and moreover $\{n_s : s \in A\}$ are linearly independent. Define $S := \mathrm{span}(\{n_s : s \in A\})$, and consider the relation

$$Rn_i = \xi_i \nu_i n_i + \sum_{\alpha=1}^{i-1} a_{\alpha i} n_\alpha.$$

for $i \in A$. This can be rewritten as

$$Rn_i = \xi_i \nu_i n_i + \sum_{\alpha \in A, \alpha \leqslant i-1} d_{\alpha i} n_\alpha, \tag{5.12}$$

for some coefficients $d_{\alpha i}$. Recall that

$$R = \sum_i k_i(k_i - 1)\lambda_i C_i^T D_i.$$

By the properties of the matrices C_i's, we see that $\mathrm{Im}(R) \subseteq S$. Now complete (if necessary) $\{n_s : s \in A\}$ to a basis $\mathcal{B}$ of $\mathbb{R}^m$ adding vectors γ_j with the property that $(\gamma_j, \gamma_k) = (\gamma_j, n_s) = 0, \, \forall \, j \neq k, s \in A$ and $\|\gamma_j\| = 1, \, \forall \, j$. By the previous observation about the image of R and (5.12), we are able to write the matrix of the linear map associated to R for the basis $\mathcal{B}$ as

$$\left[\begin{array}{ccccc|c} \xi_{i_1}\nu_{i_1} & * & * & \cdots & * & \\ 0 & \xi_{i_2}\nu_{i_2} & * & \cdots & * & \\ \vdots & \vdots & \vdots & \ddots & \vdots & \mathbf{T} \\ 0 & 0 & 0 & \cdots & \xi_{i_{\dim(S)}}\nu_{i_{\dim(S)}} & \\ \hline & \mathbf{0}_{m-\dim(S),\dim(S)} & & & \mathbf{0}_{m-\dim(S),m-\dim(S)}. \end{array} \right]$$

We denoted with $\mathbf{0}_{a,b}$ the zero matrix with a rows and b columns, with $\mathbf{T}$ the $\dim(S) \times (m-\dim(S))$ matrix of the coefficients of $R\gamma_j$ with respect to $\{n_s : s \in A\}$, and with $*$ numbers we are not interested in computing explicitly. Moreover, we chose an enumeration of A with $1 = i_1 < i_2 < \cdots < i_{\dim(S)}$. The previous matrix must have the same trace as R, so

$$0 = \mathrm{tr}(R) = \sum_{j=1}^{\dim(S)} \xi_{i_j}\nu_{i_j}.$$

6. Stationary graphs and stationary varifolds

The aim of this section is to provide the link between stationary points for energies defined on functions (or graphs) and stationary varifolds for "geometric" energies.

6.1. Notation and preliminary definitions

Recall that general m-dimensional varifolds in $\mathbb{R}^{m+n}$ (introduced by L.C. Young in [31] and pioneered in geometric measure theory by Almgren [4] and Allard [1]) are nonnegative Radon measures on the Grassmaniann of $\mathbb{G}(m, m+n)$ of (unoriented) m-dimensional planes of $\mathbb{R}^{m+n}$. In our specific case we are interested on a subclass, namely integer rectifiable varifolds, for which we can take the simpler Definition 6.1 below. A quick reference for the terminology used in this section is [7], whereas comprehensive introductions can be found in the foundational paper [1] and in the book [25].

DEFINITION 6.1. — *An integer rectifiable varifold V of dimension m is a couple (Γ, θ), where $\Gamma \subset \mathbb{R}^{m+n}$ is a m-rectifiable set in $\mathbb{R}^N$, and $\theta : \Gamma \to \mathbb{N} \setminus \{0\}$ is a Borel map.*

It is customary to denote (Γ, θ) as $\theta[\![\Gamma]\!]$ and to call θ the multiplicity of the varifold.

DEFINITION 6.2. — *Let U be an open set of $\mathbb{R}^{m+n}$, and let $\Phi : \mathbb{R}^{m+n} \to U$ be a diffeomorphism. The* pushforward *of an integer rectifiable varifold $V = \theta[\![\Gamma]\!]$ through Φ is defined as $\Phi_\# V = \theta \circ \Phi^{-1}[\![\Phi(\Gamma)]\!]$.*

For an integer rectifiable varifold $\theta[\![\Gamma]\!]$, it is customary to introduce a notion of approximate tangent plane, which exists for $\mathcal{H}^m$-a.e. point of Γ, we refer to [25, Theorem 3.1.8] for the relevant details. Provided it exists, the tangent plane at the point $y \in \Gamma$ will be denoted with $T_y\Gamma$ and it is an element of $\mathbb{G}(m, m+n)$. In the following, we will identify the Grassmanian manifold with a suitable subset of orthogonal projections, i.e. for every $L \in \mathbb{G}(m, m+n)$ we consider the linear map $P : \mathbb{R}^{m+n} \to \mathbb{R}^{m+n}$ which is the orthogonal projection onto L. With this identification we have

$$\mathbb{G}(m, m+n) \sim \left\{ P \in \mathbb{R}^{(m+n)\times(m+n)} : P = P^T, P^2 = P, \operatorname{rank}(P) = \operatorname{tr}(P) = m \right\}.$$

We are interested in graphs of maps $u : \Omega \subset \mathbb{R}^m \to \mathbb{R}^n$, and we always consider $\mathbb{R}^m = \operatorname{span}\{e_1, \ldots, e_m\}$, where $\{e_1, \ldots, e_{n+m}\}$ is the canonical basis of $\mathbb{R}^{n+m}$. In other words, we are interested in sets of the form $\Gamma_u = \{z \in \mathbb{R}^{n+m} : z = (x, u(x)), x \in \Omega\}$. For this reason, we need to characterize the

space of orthogonal projections on tangent planes to graphs (on the plane $\text{span}\{e_1, \ldots, e_m\}$). Since at the differentiability point x_0, we have

$$D((x, u(x)))|_{x=x_0} = \begin{pmatrix} \text{id}_m \\ Du(x_0) \end{pmatrix} \in \mathbb{R}^{(m+n) \times m},$$

it is convenient to introduce the following notation:

$$M(X) := \begin{pmatrix} \text{id}_m \\ X \end{pmatrix}.$$

Therefore, every tangent plane to a graph Γ_u is of the form

$$\tau(X) = \text{span}\{M(X)^T e_1, \ldots, M(X)^T e_{n+m}\}.$$

With the notation above, the tangent plane of Γ_u at x_0 is $\tau(Du(x_0))$. The orthogonal projection on $\tau(X)$ is given by the formula

$$h(X) := M(X)S(X)M(X)^T$$

where

$$S(X) := (M(X)^T M(X))^{-1},$$

or, more explicitly,

$$h(X) = \left[\begin{array}{c|c} h_1(X) & h_3(X) \\ \hline h_2(X) & h_4(X) \end{array} \right] = \left[\begin{array}{c|c} S(X) & S(X)X^T \\ \hline XS(X) & XS(X)X^T \end{array} \right]. \tag{6.1}$$

In particular, using the notation above, we remark that $T_{x_0}\Gamma_u = h(Du(x_0))$. This discussion motivates the following

DEFINITION 6.3. — *We denote by* $\mathbb{G}_0(m, m+n) := h(\mathbb{R}^{n \times m}) \subset \mathbb{G}(m, m+n)$ *the set of orthogonal projections of tangent planes to graphs of maps defined on* $\text{span}\{e_1, \ldots, e_m\} \subset \mathbb{R}^{n+m}$.

We will use in general, i.e. for any matrix $M \in \mathbb{R}^{(m+n) \times (m+n)}$ the same splitting as in (6.1):

$$M = \left[\begin{array}{c|c} M_1 & M_3 \\ \hline M_2 & M_4 \end{array} \right] \tag{6.2}$$

with $M_1 \in \mathbb{R}^{m \times m}$, $M_4 \in \mathbb{R}^{n \times n}$. Using this notation, it is not difficult to verify that

$$h^{-1}(P) = P_2 P_1^{-1}. \tag{6.3}$$

The map h is therefore a smooth diffeomorphism between $\mathbb{R}^{n \times m}$ and the open subset $\mathbb{G}_0$.

In this section, we will use freely the following fact. Recall that, by (1.4), for every $X \in \mathbb{R}^{n \times m}$ the area element is given by

$$\mathcal{A}(X) = \sqrt{\det(\text{id}_m + X^T X)}.$$

By the Cauchy–Binet formula, [6, Proposition 2.69],

$$\mathcal{A}(X) = \sqrt{1 + \|X\|^2 + \sum_{r=2}^{\min\{m,n\}} \sum_{Z \in \mathcal{A}_r} \det(X^Z)^2},$$

where we used the notation introduced in Definition 3.5.

Finally, throughout the section, we use the following notation:

- if $z \in \mathbb{R}^m \times \mathbb{R}^n$, then we will write $z = (x, y), x \in \mathbb{R}^m, y \in \mathbb{R}^n$;
- $\pi : \mathbb{R}^m \times \mathbb{R}^n \to \mathbb{R}^m$ denotes the projection on the first factor, i.e. $\pi(z) = \pi((x, y)) = x$.

6.2. Graphs and varifolds

If $u \in W^{1,p}(\Omega, \mathbb{R}^n)$, $\Omega \subset \mathbb{R}^m$ and $p > m$, Morrey's embedding theorem shows the existence of a precise representative of u which is Hölder continuous. In what follows we will always assume that the map u is given pointwise by such (Hölder) continuous precise representative. As done above, we use the notation Γ_u for the (set-theoretic) graph $\{(x, u(x)) : x \in \Omega\}$, which is a relatively closed subset of $\Omega \times \mathbb{R}^n$. The classical area formula (see for instance [16, Corollary 2, Chapter 3]) implies that Γ_u is m-rectifiable and its $\mathcal{H}^m$ measure is given by

$$\int_\Omega \mathcal{A}(Du).$$

We can thus consider the corresponding varifold $[\![\Gamma_u]\!]$.

If $u \in W^{1,m}(\Omega, \mathbb{R}^n)$, then u has a precise representative which is however defined only up to a set of m-capacity 0 (but not *everywhere*). Moreover, if for maps $u \in W^{1,m} \cap C(\Omega, \mathbb{R}^n)$, for which the set-theoretic graph Γ_u could be defined classically, it can be proven that Γ_u does not necessarily have locally finite $\mathcal{H}^m$-measure, in spite of the fact that $\mathcal{A}(Du)$ belongs to L^1_{loc}. In particular the area formula fails. For this reason, following the notation and terminology of [16, Section 1.5, 2.1], we introduce the "rectifiable part of the graph of u", which will be denoted by $\mathcal{G}_u$ (the notation in [16] is in fact $\mathcal{G}_{u,\Omega}$: we will omit the domain Ω since in our case it is always clear from the context).

First we denote the set of Lebesgue points of u by $\mathcal{L}_u$ and we introduce the set

$$A_D(u) := \{x \in \Omega : u \text{ is approximately differentiable at } x\}.$$

For the definition of approximate differentiability, see [16, Section 1.4, Definition 3]. We also set

$$\mathcal{R}_u := A_D(u) \cap \mathcal{L}_u.$$

Notice that, since $u \in W^{1,m}(\Omega, \mathbb{R}^n)$, then $|\Omega \setminus \mathcal{R}_u| = 0$. From now on, we always assume that u so that $u(x)$ is the Lebesgue value at every point $x \in \mathcal{R}_u$. The *rectifiable part of the graph* of u is then

$$\mathcal{G}_u := \{(x, u(x)) : x \in \mathcal{R}_u\}.$$

By [16, Section 1.5, Theorem 4], $\mathcal{G}_u$ is m-rectifiable and

$$\mathcal{H}^m(\mathcal{G}_u) = \int_\Omega \mathcal{A}(Du(x)) \, \mathrm{d}x.$$

Since $\mathcal{A}(Du) \in L^1_{loc}$, this allows us to introduce the integer rectifiable varifold[4] $[\![\mathcal{G}_u]\!]$. When $u \in W^{1,p}$ for $p > m$, the Lusin property (namely the fact that $v(x) := (x, u(x))$ maps sets of Lebesgue measure zero in sets of $\mathcal{H}^m$-measure zero, cf. again [16]) and Morrey's embedding imply $\mathcal{G}_u \subset \Gamma_u$ and $\mathcal{H}^m(\Gamma_u \setminus \mathcal{G}_u) = 0$. In particular $[\![\mathcal{G}_u]\!] = [\![\Gamma_u]\!]$.

By [16, Section 1.5, Theorem 5], the approximate tangent plane $T_y \mathcal{G}_u$ coincides for $\mathcal{H}^m$-a.e. $z_0 = (x_0, u(x_0)) \in \mathcal{G}_u$ or, with

$$T_z \mathcal{G}_u = \{(x, Du(x_0)x) : x \in \mathbb{R}^m\} \in \mathbb{G}_0(m, m+n).$$

The following proposition allows then to pass from functionals defined on varifolds to classical functionals in the vectorial calculus of variations (and viceversa).

PROPOSITION 6.4. — *Let $u \in W^{1,m}(\Omega, \mathbb{R}^n)$, and define $v(x) := (x, u(x))$. Denote with $C_b(\Omega \times \mathbb{R}^n \times \mathbb{G}_0)$ the space of continuous and bounded functions on $\Omega \times \mathbb{R}^n \times \mathbb{G}_0$. Then, for every $\varphi \in C_b(\Omega \times \mathbb{R}^n \times \mathbb{G}_0)$, the following holds*

$$[\![\mathcal{G}_u]\!](\varphi) := \int_{\mathcal{G}_u} \varphi(z, T_z \mathcal{G}_u) \, \mathrm{d}\mathcal{H}^m(z)$$

$$= \int_\Omega \varphi(v(x), h(Du(x))) \mathcal{A}(Du(x)) \, \mathrm{d}x. \quad (6.4)$$

Consider therefore a functional

$$\mathbb{E}(u) := \int_\Omega f(Du(x)) \, \mathrm{d}x,$$

for some $f : \mathbb{R}^{n \times m} \to \mathbb{R}$ with

$$\frac{f(X)}{\mathcal{A}(X)} \in C_b(\mathbb{R}^{n \times m}).$$

---

[4] In fact the $\mathcal{G}_u$ can be oriented to give an integer rectifiable current of multiplicity 1 and without boundary in $\Omega \times \mathbb{R}^n$, see [16, Proposition 1, Section 2.1]. The varifold that we consider is then the one induced by the current in the usual sense.

Define moreover $F, G : \mathbb{G}_0 \to \mathbb{R}$ as

$$F(M) := f(h^{-1}(M)), \; G(M) := \mathcal{A}(h^{-1}(M)).$$

For any map $u \in W^{1,m}(\Omega, \mathbb{R}^n)$, we can apply (6.4) to write:

$$\int_\Omega f(Du(x)) \, \mathrm{d}x = \int_\Omega F(h(Du(x))) \, \mathrm{d}x = \int_\Omega \frac{F(h(Du(x)))}{G(h(Du(x)))} \mathcal{A}(Du(x)) \, \mathrm{d}x$$
$$= \int_{\mathcal{G}_u} \Psi(T_z \mathcal{G}_u) \, \mathrm{d}\mathcal{H}^m(z),$$

where we have defined the map Ψ on the open subset $\mathbb{G}_0$ of the Grassmanian $\mathbb{G}(m, m+n)$ as

$$\Psi(h(X)) := \frac{F(h(X))}{G(h(X))} = \frac{f(X)}{\mathcal{A}(X)}. \tag{6.5}$$

We are thus ready to introduce the following functional

DEFINITION 6.5. — *Let $V = \theta[\![\Gamma]\!]$ be an m-dimensional integer rectifiable varifold in $\mathbb{R}^{m+n}$ with the property that the approximate tangent $T_x\Gamma$ belongs to $\mathbb{G}_0$ for $\mathcal{H}^m$-a.e. $x \in \Gamma$. Then*

$$\Sigma(V) = \int_\Gamma \Psi(T_x\Gamma)\theta(x) \, \mathrm{d}\mathcal{H}^m(x).$$

The above discussion then proves the following

PROPOSITION 6.6. — *If $\Omega \subset \mathbb{R}^m$ and $u \in W^{1,m}(\Omega, \mathbb{R}^n)$, then $\Sigma([\![\mathcal{G}_u]\!]) = \mathbb{E}(u)$. Moreover, if $u \in W^{1,p}(\Omega, \mathbb{R}^n)$ with $p > m$, then $\Sigma([\![\Gamma_u]\!]) = \mathbb{E}(u)$.*

6.3. First variations

We do not address here the issue of extending the functional Σ to general varifolds (namely of extending Ψ to all of $\mathbb{G}(m, m+n)$). Rather, assuming that such an extension exists, we wish to show that the usual stationarity of varifolds with respect to the functional Σ is equivalent to stationarity with respect to two particular classes of deformations, which reduce to inner and outer variations in the case of graphs. We start recalling the usual stationarity condition.

DEFINITION 6.7. — *Let $\Psi : \mathbb{G}(m, m+n) \to [0, \infty]$ be a continuous function. Fix a vector field $g \in C_c^1(\mathbb{R}^{m+n}; \mathbb{R}^{m+n})$ and define X_ε as the flow generated by g, namely $X_\varepsilon(x) = \gamma_x(\varepsilon)$, if γ_x is the solution of the following system*

$$\begin{cases} \gamma'(t) = g(\gamma(t)) \\ \gamma(0) = x. \end{cases}$$

We define the variation of V with respect to the vector field $g \in C_c^1(\mathbb{R}^{m+n};$ $\mathbb{R}^{m+n})$ as

$$[\delta_\Psi V](g) := \lim_{\varepsilon \to 0} \frac{\Sigma((X_\varepsilon)_\# V) - \Sigma(V)}{\varepsilon}.$$

V is said to be stationary *if $[\delta_\Psi V](g) = 0, \forall\, g \in C_c^1(\mathbb{R}^{m+n}; \mathbb{R}^{m+n})$.*

Recall that the left hand side of (1.2) is the derivative of the energy along the variation $u(x) + \varepsilon v(x)$. The latter corresponds, at the infinitesimal level, to the one-parameter family of deformations of the graph induced by the vector field $g(x, y) = (0, v(x))$. Similarly, (1.3) is given differentiating the variation $u \circ X_\varepsilon$ where $X_\varepsilon(x) = x + \varepsilon\Phi(x)$, which corresponds to the one-parameter family of deformations of the graph induced by the vector field $(-\Phi(x), 0)$. These remarks can be used in order to show rigorously that, if $[\![\mathcal{G}_u]\!]$ is stationary in the sense of varifolds (for the energy corresponding to $\mathbb{E}$), then u satisfies (1.2) and (1.3). The converse is less obvious: even though any vector field $g(x, y)$ can be decomposed into a horizontal and vertical part $(g_1(x, y), 0) + (0, g_2(x, y))$, there is still the issue that the g_i's depend on the variable y as well. When the graph u is smooth, we can simply argue that variations of the graph along the vector field $g(x, y)$ are equal to variations along $\widetilde{g}(x) := g(x, u(x))$. This however creates several technical difficulties if we only assume Sobolev regularity for u. Nonetheless the conclusion is still correct. We conclude this section with a rather general equivalence statement between stationarity of graphs and stationarity of varifolds, for which we need first some suitable terminology and notation. The (somewhat lengthy) proof is postponed to the next section.

Given an orthogonal projection $P \in \mathbb{G}_{m,m+n)}$, we denote $P^\perp := \mathrm{id}_{m+n} - P$. The notation $P^\perp$ is due to the fact that, if P represents the orthogonal projection onto the m-plane $\tau \subset \mathbb{R}^{n+m}$, $\mathrm{id}_{n+m} - P$ is the element in $\mathbb{G}(n, m+n)$ representing the orthogonal projection onto the n-plane $\tau^\perp \subset \mathbb{R}^{n+m}$. From [9, Lemma A.2], we know that, for $V = \theta[\![\Gamma]\!]$,

$$[\delta_\Psi(V)](g) = \int_\Gamma \langle B_\Psi(T_x\Gamma), Dg(x)\rangle \theta(x)\, \mathrm{d}\mathcal{H}^m(x),$$

$$\forall\, g \in C_c^1(\mathbb{R}^{m+n}, \mathbb{R}^{m+n}) \quad (6.6)$$

where $B_\Psi(\,\cdot\,) : \mathbb{G}(m, m+n) \to \mathbb{R}^{(m+n)\times(m+n)}$ is defined through the relation

$$\langle B_\Psi(P), L\rangle = \Psi(P)\langle P, L\rangle + \langle d\Psi(P), P^\perp LP + (P^\perp LP)^T\rangle,$$

$$\forall\, P \in \mathbb{G}(m, m+n), \forall\, L \in \mathbb{R}^{(m+n)\times(m+n)}, \quad (6.7)$$

We are now ready to state our desired equivalence between stationarity of the map u for the energy $\mathbb{E}$ and stationarity of the varifold $[\![\mathcal{G}_u]\!]$ for the

corresponding functional Σ. In what follows, given a function g on $\mathcal{G}_u$ we will use the shorthand notation $\|g\|_{q,\mathcal{G}_u}$ for the norm $\|g\|_{L^q(\mathcal{H}^m \llcorner \mathcal{G}_u)}$.

PROPOSITION 6.8. — *Assume that $f \in C^1(\mathbb{R}^{n \times m})$ admits an extension $\Psi \in C^1(\mathbb{G}(m, m+n))$, in the sense that (6.5) holds for every $X \in \mathbb{R}^{n \times m}$. Fix any $m \leqslant p \leqslant +\infty$, $1 \leqslant q < +\infty$ and a Lipschitz, bounded, open set $\Omega \subset \mathbb{R}^m$. If a map $u \in W^{1,p}(\Omega, \mathbb{R}^n)$ satisfies*

$$\begin{cases} \left| \int_\Omega \langle Df(Du), Dv \rangle \, \mathrm{d}x \right| \leqslant C \|v\mathcal{A}^{\frac{1}{q}}(Du)\|_q & \forall \, v \in C_c^1(\Omega, \mathbb{R}^n) \\[2mm] \left| \int_\Omega \langle Df(Du), DuD\Phi(x) \rangle \mathrm{d}x - \int_\Omega f(Du)\mathrm{div}(\Phi)\mathrm{d}x \right| \leqslant C \|\Phi \mathcal{A}^{\frac{1}{q}}(Du)\|_q \\[2mm] & \forall \, \Phi \in C_c^1(\Omega, \mathbb{R}^m), \end{cases} \tag{6.8}$$

for some $C \geqslant 0$, then the integer rectifiable varifold $[\![\mathcal{G}_u]\!]$ in $\mathbb{R}^{m+n}$ satisfies

$$|\delta_\Psi([\![\mathcal{G}_u]\!])(g)| \leqslant C' \|g\|_{q,\mathcal{G}_u} \qquad \forall \, g \in C_c^1(\Omega \times \mathbb{R}^n, \mathbb{R}^{m+n}), \tag{6.9}$$

for some number $C' = C'(C, m, p, q) \geqslant 0$. Conversely, if (6.9) holds for some C', then (6.8) holds for some $C = C(C', m, p, q)$. Moreover, $C' = 0$ if and only $C = 0$, namely u is stationary for the energy $\mathbb{E}$ if and only if $[\![\mathcal{G}_u]\!]$ is stationary for the energy Σ.

Remark 6.9. — As already noticed, when $p > m$ we can replace $[\![\mathcal{G}_u]\!]$ with $[\![\Gamma_u]\!]$. Moreover, under such stronger assumption, the proposition holds also for $q = \infty$, provided we set $\mathcal{A}(Du)^{\frac{1}{q}} := 1$ in that case. Finally, if $p = \infty$, then we can drop the request that f admits a C^1 extension Ψ, and the same proof would work if we extended Ψ as in (6.5) as $\Psi(T) \equiv +\infty$, if $T \notin \mathbb{G}_0(m, m+n)$.

The proof of the previous proposition is a consequence of a few technical lemmas and will be given in the next section.

7. Proof of Proposition 6.8

Let $f \in C^1(\mathbb{R}^{n \times m})$ be of the form $f(X) = \Psi(h(X))\mathcal{A}(X)$. In the next lemma we study the growth of the matrix-fields associated to the inner and the outer variations, i.e.

$$A(X) := Df(X) \tag{7.1}$$

$$B(X) := f(X)\,\mathrm{id}_m - X^T Df(X). \tag{7.2}$$

Define also the matrix-field $V_f : \mathbb{R}^{n \times m} \to \mathbb{R}^{(m+n) \times (m+n)}$ to be

$$V_f(X) := \frac{1}{\mathcal{A}(X)} \left[\begin{array}{c|c} B(X) & B(X)X^T \\ \hline A(X) & A(X)X^T \end{array} \right]. \tag{7.3}$$

In Lemma 7.2, we will prove that

$$B_\Psi(h(X)) = V_f(X), \quad \forall\, X \in \mathbb{R}^{n \times m}.$$

Combining Lemma 7.1 and 7.2 with the area formula we obtain Lemma 7.3, from which we will infer Proposition 6.8.

LEMMA 7.1. — *Let $\Psi \in C^1(\mathbb{G}(m, m+n))$ and let $f(X) = \Psi(h(X))\mathcal{A}(X)$, where h is the map defined in* (6.1). *Then,*

$$\|A(X)\| \lesssim 1 + \|X\|^{\min\{m,n\}-1}, \|B(X)\| \lesssim 1 + \|X\|^{\min\{n,m-1\}}. \quad (7.4)$$

In the statement of the Lemma and in the proof, the symbol $\Lambda \lesssim \Xi$ means that there exist a non-negative constant C depending only on n, m and on $\|\Psi\|_{C^1(\mathbb{G}(m,m+n)}$ such that

$$\Lambda \leqslant C\,\Xi.$$

The lemma above is needed to get reach enough summability in order to justify the integral formulas in (the statement and the proof of) Lemma 7.3. In some sense it is thus less crucial than the next lemma, which contains instead the core computations. For these reasons, the argument of Lemma 7.1, which contains several lengthy computations is given in the appendix.

LEMMA 7.2. — *For every $X \in \mathbb{R}^{n \times m}$,*

$$B_\Psi(h(X)) = V_f(X).$$

LEMMA 7.3. — *Let $f(X) = \Psi(h(X))\mathcal{A}(X)$ be a function of class $C^1(\mathbb{R}^{n \times m})$. Then, for every $g = (g^1, \ldots, g^{m+n}) \in C_c^1(\Omega \times \mathbb{R}^n)$, the following equality holds:*

$$\delta_\Psi([\![\mathcal{G}_u]\!])(g) = \int_\Omega \langle B(Du(x)), D(g_1(x, u(x)))\rangle \, \mathrm{d}x$$

$$+ \int_\Omega \langle A(Du(x)), D(g_2(x, u(x)))\rangle \, \mathrm{d}x, \quad (7.5)$$

where

$$g_1(x, y) := (g^1(x, y), \ldots, g^m(x, y)),$$
$$g_2(x, y) := (g^{m+1}(x, y), \ldots, g^{m+n}(x, y))$$

and $A(X)$ and $B(X)$ are as in (7.1) *and* (7.2).

We next prove Lemma 7.2 and Lemma 7.3 and hence end the section showing how to use Lemma 7.3 to conclude the desired Proposition 6.8.

7.1. Proof of Lemma 7.2

For a map $g : \mathbb{G}(m + n, m) \to \mathbb{R}^\ell$, $\ell \geqslant 1$, of class C^1, we denote the differential at the point $P \in \mathbb{G}(m + n, m)$ with the symbol $d_P g$. For $H \in T_P \mathbb{G}(m+n, m)$, and for $\gamma : (-1, 1) \to \mathbb{G}(m+n, m)$ with $\gamma(0) = P$, $\gamma'(0) = H$, we denote

$$d_P g(P)[H] := \lim_{t \to 0} \frac{g(\gamma(t)) - g(P)}{t}.$$

If $\ell = 1$, we identify $d_P g(P)$ with the $\mathbb{R}^{(m+n) \times (m+n)}$ associated matrix representing the differential, and we denote $d_P g(P)[H]$ with $\langle d_P g(P), H \rangle$. Moreover, we recall the splitting introduced in (6.2), namely for any matrix $M \in \mathbb{R}^{(m+n) \times (m+n)}$ we denote

$$M = \left[\begin{array}{c|c} M_1 & M_3 \\ \hline M_2 & M_4 \end{array} \right]$$

with $M_1 \in \mathbb{R}^{m \times m}$, $M_4 \in \mathbb{R}^{n \times n}$. In this proof, we will use the following facts:

- The tangent plane of $\mathbb{G}(m, m + n)$ at the point P is given by

$$T_P \mathbb{G}(m, m + n) = \left\{ M \in \mathbb{R}^{(m+n) \times (m+n)} : \begin{array}{l} M = P^\perp L P + (P^\perp L P)^T, \\ \text{for some } L \in \mathbb{R}^{(m+n) \times (m+n)} \end{array} \right\},$$

  as proved in [9, Appendix A].
- Let $h : \mathbb{R}^{n \times m} \to \mathbb{G}_0$ be the map defined in (6.1). Recall that its inverse is given by $h^{-1}(P) = P_2 P_1^{-1}$. For every $H \in T_P \mathbb{G}(m, m+n)$, one has:

$$d_P(h^{-1})(P)[H] = (H_2 - P_2 P_1^{-1} H_1) P_1^{-1} \in \mathbb{R}^{n \times m}. \tag{7.6}$$

- Recall that the area functional is defined as

$$\mathcal{A}(X) = \sqrt{\det(M(X)^T M(X))} \qquad \text{where } M(X) = \left[\begin{array}{c} \mathrm{id}_m \\ X \end{array} \right].$$

  Hence, for every $X, Y \in \mathbb{R}^{n \times m}$, we have

$$\langle D\mathcal{A}(X), Y \rangle = \frac{1}{2} \mathcal{A}(X) \operatorname{tr}[(M(X)^T M(X))^{-1} (Y^T X + X^T Y)]. \tag{7.7}$$

Recall the definition of $B_\Psi(P)$ given in (6.7). Since

$$\Psi(P) = \frac{f(h^{-1}(P))}{\mathcal{A}(h^{-1}(P))},$$

for every $H \in T_P \mathbb{G}(m, m+n)$ we have

$$\langle d_P \Psi(P), H \rangle = \frac{1}{\mathcal{A}(X)} \langle Df(h^{-1}(P)), d_P(h^{-1})(P)[H] \rangle$$
$$- \frac{f(h^{-1}(P))}{\mathcal{A}^2(h^{-1}(P))} \langle D\mathcal{A}(h^{-1}(P)), d_P(h^{-1})(P)[H] \rangle.$$

When evaluated at $P = h(X)$, the previous expression reads

$$\langle d_P \Psi(h(X)), H \rangle = \frac{1}{\mathcal{A}(X)} \langle Df(X), d_P(h^{-1})(h(X))[H] \rangle$$
$$- \frac{f(X)}{\mathcal{A}^2(X)} \langle D\mathcal{A}(X), d_P(h^{-1})(h(X))[H] \rangle. \quad (7.8)$$

By (6.7), we know that, for every $L \in \mathbb{R}^{(m+n) \times (m+n)}$,

$$\langle B_\Psi(h(X)), L \rangle = \Psi(h(X)) \langle h(X), L \rangle$$
$$+ \langle d_P \Psi(h(X)), h(X)^\perp L h(X) + (h(X)^\perp L h(X))^T \rangle.$$

Therefore, we want to compute (7.8) when

$$H = h(X)^\perp L h(X) + (h(X)^\perp L h(X))^T$$
$$= Lh(X) - h(X)Lh(X) + h(X)L^T - h(X)L^T h(X).$$

We wish to find an expression for

$$d_P(h^{-1})(h(X))[h(X)^\perp L h(X) + h(X)L^T h(X)^\perp].$$

Using the decomposition introduced in (6.2) of L in 4 submatrices, we compute

$$Lh(X) = \left[\begin{array}{c|c} L_1 & L_3 \\ \hline L_2 & L_4 \end{array} \right] \left[\begin{array}{c|c} S & SX^T \\ \hline XS & XSX^T \end{array} \right]$$
$$= \left[\begin{array}{c|c} L_1 S + L_3 XS & L_1 SX^T + L_3 XSX^T \\ \hline L_2 S + L_4 XS & L_2 SX^T + L_4 XSX^T \end{array} \right] \quad (7.9)$$

and

$$h(X)Lh(X)$$
$$= \left[\begin{array}{c|c} S(L_1+L_3X+X^TL_2+X^TL_4X)S & S(L_1+L_3X+X^TL_2+X^TL_4X)SX^T \\ \hline XS(L_1+L_3X+X^TL_2+X^TL_4X)S & XS(L_1+L_3X+X^TL_2+X^TL_4X)SX^T \end{array} \right] \quad (7.10)$$

Combining (7.6) with (7.10), we get

$$d_P(h^{-1})(h(X))[Lh(X)]$$
$$= (L_2 S + L_4 XS - XSS^{-1}L_1 S - XSS^{-1}L_3 XS)S^{-1} \quad (7.11)$$
$$= L_2 + L_4 X - XL_1 - XL_3 X,$$

$$d_P(h^{-1})(h(X))[h(X)Lh(X)]$$
$$= XS(L_1 + L_3X + X^T L_2 + X^T L_4 X - S^{-1}SL_1 - S^{-1}SL_3X$$
$$\quad - S^{-1}SX^T L_2 - S^{-1}SX^T L_4 X)SS^{-1} \qquad (7.12)$$
$$= XS(L_1 + L_3X + X^T L_2 + X^T L_4 X - L_1 - L_3X - X^T L_2 - X^T L_4 X)$$
$$= 0$$

and

$$d_P(h^{-1})(h(X))[h(X)L^T]$$
$$\qquad = d_P(h^{-1})(h(X))[(L \circ h(X))^T] \qquad (7.13)$$
$$\qquad = (XSL_1^T + XSX^T L_3^T - XSL_1^T - XSX^T L_3^T)S^{-1} = 0.$$

Combining (7.11), (7.12) and (7.13), we get that

$$d_P(h^{-1})(h(X))[h(X)^\perp Lh(X) + h(X)L^T h(X)^\perp]$$
$$\qquad\qquad\qquad\qquad = d_P(h^{-1})(h(X))(Lh(X))$$
$$\qquad\qquad\qquad\qquad = L_2 + L_4X - XL_1 - XL_3X.$$

Now define the matrix:

$$C := L_2 + L_4X - XL_1 - XL_3X.$$

To expand (7.8), we now need to rewrite

$$\langle D\mathcal{A}(X), d_P(h^{-1})(h(X))[H] \rangle.$$

First, we must compute the trace part coming from (7.7):

$$\mathrm{tr}[S(C^T X + X^T C)]$$
$$\quad = \mathrm{tr}[S(L_2^T X + X^T L_4^T X - L_1^T X^T X - X^T L_3^T X^T X)]$$
$$\quad\quad + \mathrm{tr}[S(X^T L_2 + X^T L_4 X - X^T XL_1 - X^T XL_3X)]$$
$$\quad = 2\,\mathrm{tr}(SX^T L_2) + 2\,\mathrm{tr}(SX^T L_4 X) - 2\,\mathrm{tr}(SX^T XL_1) - 2\,\mathrm{tr}(SX^T XL_3X).$$

Hence, if $H = h(X)^\perp Lh(X) + h(X)L^T h(X)^\perp$, we have just proved that:

$$\langle d_P \Psi(h(X)), H \rangle = \frac{1}{\mathcal{A}(X)} \langle Df(X), L_2 + L_4X - XL_1 - XL_3X \rangle$$
$$- \frac{f(X)}{\mathcal{A}(X)}(\mathrm{tr}(SX^T L_2) + \mathrm{tr}(SX^T L_4 X) - \mathrm{tr}(SX^T XL_1) - \mathrm{tr}(SX^T XL_3X)). \qquad (7.14)$$

To conclude, we also need to compute

$$\Psi(h(X))\langle h(X), L \rangle$$
$$\quad = \frac{f(X)}{\mathcal{A}(X)}(\langle L_1, S \rangle + \langle L_2, XS \rangle + \langle L_3, SX^T \rangle + \langle L_4, XSX^T \rangle) \qquad (7.15)$$
$$\quad = \frac{f(X)}{\mathcal{A}(X)}(\mathrm{tr}(SL_1) + \mathrm{tr}(SX^T L_2) + \mathrm{tr}(XSL_3) + \mathrm{tr}(XSX^T L_4)).$$

Now we sum (7.14) and (7.15) to get $\langle B_\Psi(h(X)), L\rangle$. Using that $S^{-1}(X) = X^T X + \mathrm{id}_m$ and the invariance of the trace under cyclic permutations, we rewrite

$$\mathrm{tr}(SL_1) + \mathrm{tr}(SX^T L_2) + \mathrm{tr}(XSL_3) + \mathrm{tr}(XSX^T L_4) - \mathrm{tr}(SX^T L_2)$$
$$- \mathrm{tr}(SX^T L_4 X) + \mathrm{tr}(SX^T X L_1) + \mathrm{tr}(SX^T X L_3 X) = \mathrm{tr}(L_1) + \mathrm{tr}(L_3 X).$$

Combining our previous computations, we find

$$\langle B_\Psi(h(X)), L\rangle$$
$$= \frac{f(X)}{\mathcal{A}(X)}(\mathrm{tr}(L_1) + \mathrm{tr}(L_3 X)) + \frac{1}{\mathcal{A}(X)}\langle Df(X), L_2 + L_4 X - X L_1 - X L_3 X\rangle$$
$$= \frac{1}{\mathcal{A}(X)}[-\langle X^T Df(X) + f(X)\,\mathrm{id}_m, L_1\rangle + \langle Df(X), L_2\rangle$$
$$+ \langle f(X) X^T - X^T Df(X) X^T, L_3\rangle + \langle Df(X) X^T, L_4\rangle].$$

Since L was arbitrary, we conclude that

$$B_\Psi(h(X)) = \frac{1}{\mathcal{A}(X)}\left[\begin{array}{c|c} B(X) & B(X)X^T \\ \hline A(X) & A(X)X^T \end{array}\right].$$

7.2. Proof of Lemma 7.3

Fix g as in the statement of the Lemma. By (6.6), we know that

$$\delta_\Psi(\llbracket \mathcal{G}_u \rrbracket)(g) = \int_{\mathcal{G}_u} \langle B_\Psi(T_z \mathcal{G}_u), Dg(z)\rangle\,\mathrm{d}\mathcal{H}^m(z).$$

Now define

$$F(z, T) := \langle B_\Psi(T), Dg(z)\rangle$$
$$\text{and} \quad \overline{F}(x, u(x)) := \langle B_\Psi(h(Du(x))), Dg(x, u(x))\rangle.$$

We have $F \in C_c(\Omega \times \mathbb{R}^n \times \mathbb{G}(m, m+n))$ and we apply Proposition 6.4 to find the equality

$$\int_{\mathcal{G}_u} \langle B_\Psi(T_z \Gamma_u), Dg(z)\rangle\,\mathrm{d}\mathcal{H}^m(z) = \int_\Omega \mathcal{A}(Du(x))\overline{F}(x, u(x))\,\mathrm{d}x. \qquad (7.16)$$

By Lemma 7.2,

$$\overline{F}(x, u(x)) = \langle V_f(Du(x)), Dg(x, u(x))\rangle$$

a.e. in Ω. Moreover, since

$$\mathcal{A}(Du(x))V_f(Du(x)) = \left[\begin{array}{c|c} B(Du(x)) & B(Du(x))Du(x)^T \\ \hline A(Du(x)) & A(Du(x))Du(x)^T \end{array}\right],$$

we have

$$\mathcal{A}(Du(x))\overline{F}(x, u(x))$$
$$= \langle D_x g^1(x, u(x)), B(Du(x))\rangle + \langle B(Du(x))Du^T(x), D_y g^1(x, u(x)))\rangle$$
$$+ \langle D_x g^2(x, u(x)), A(Du(x))\rangle + \langle A(Du(x))Du^T(x), D_y g^2(x, u(x)))\rangle$$
$$= \langle B(Du(x)), D(g^1(x, u(x)))\rangle + \langle A(Du(x)), D(g^2(x, u(x)))\rangle.$$

The previous equality and (7.16) yield the conclusion. $\blacksquare$

7.3. Proof of Proposition 6.8

First, assume (6.8), and fix any $g \in C_c^1(\Omega \times \mathbb{R}^n, \mathbb{R}^{m+n})$, $g = (g^1, \dots, g^{m+n})$. Define

$$\overline{\Phi}(x) := (g^1(x, u(x)), \dots, g^m(x, u(x)))$$
$$\bar{v}(x) := (g^{m+1}(x, u(x)), \dots, g^{m+n}(x, u(x))).$$

We have $\overline{\Phi} \in L^\infty \cap W_0^{1,m}(\Omega, \mathbb{R}^m)$ and $\bar{v} \in L^\infty \cap W_0^{1,m}(\Omega, \mathbb{R}^n)$. Notice that we require (6.8) to hold only for C^1 maps with compact support, but Lemma 7.1 implies through an approximation argument that

$$\begin{cases} \left| \int_\Omega A(Du), Dv \rangle \, \mathrm{d}x \right| \leqslant C \|v \mathcal{A}^{\frac{1}{q}}(Du)\|_q, \quad \forall\, v \in L^\infty \cap W_0^{1,m}(\Omega, \mathbb{R}^n) \\ \left| \int_\Omega B(Du), D\Phi \rangle \, \mathrm{d}x \right| \leqslant C \|\Phi \mathcal{A}^{\frac{1}{q}}(Du)\|_q, \quad \forall\, \Phi \in L^\infty \cap W_0^{1,m}(\Omega, \mathbb{R}^m). \end{cases} \quad (7.17)$$

Indeed, to prove, for instance, that the first inequality holds for any $v \in L^\infty \cap W_0^{1,m}$, pick a sequence $v_k \in C_c^\infty(\Omega, \mathbb{R}^n)$ such that $\|v_k\|_{L^\infty}$ is equibounded and $v_k \to v$ in $W^{1,m}$, L^q and pointwise a.e.. The fact that

$$\int_\Omega \langle A(Du), Dv_k \rangle \, \mathrm{d}x \longrightarrow \int_\Omega \langle A(Du), Dv \rangle \, \mathrm{d}x$$

is an easy consequence of the $W^{1,m}$ convergence of v_k to v and the fact that $A(Du) \in W^{\frac{m}{m-1}}(\Omega, \mathbb{R}^{n \times m})$ by Lemma 7.1. Moreover, the quantity

$$\|v_k \mathcal{A}^{\frac{1}{q}}(Du)\|_q \longrightarrow \|v \mathcal{A}^{\frac{1}{q}}(Du)\|_q$$

by the dominated convergence theorem. Indeed, we required the pointwise convergence of v_k to v and moreover we can bound for every k and almost every $x \in \Omega$:

$$\|v_k \mathcal{A}^{\frac{1}{q}}(Du)\|^q(x) \leqslant \sup_k \|v_k\|_{L^\infty}^q \mathcal{A}(Du(x)) \in L^1(\Omega).$$

Hence (7.17) with v_k instead of v implies the same inequality for v by taking the limit as $k \to \infty$. The proof of the second inequality of (7.17) is analogous. We combine (7.17) with (7.5) to write

$$|\delta_\Psi([\![\mathcal{G}_u]\!])(g)| \leqslant \left| \iint_\Omega \langle A(Du), D\bar{v} \rangle \, \mathrm{d}x \right| + \left| \iint_\Omega \langle B(Du), D\bar{\Phi} \rangle \, \mathrm{d}x \right|$$
$$\leqslant C(\|\bar{v}\mathcal{A}^{\frac{1}{q}}(Du)\|_q + \|\bar{\Phi}\mathcal{A}^{\frac{1}{q}}(Du)\|_q).$$

Notice that, since $\bar{v}(\,\cdot\,, u(\,\cdot\,)) \in L^\infty(\Omega, \mathbb{R}^n)$ and $\bar{\Phi}(\,\cdot\,, u(\,\cdot\,)) \in L^\infty(\Omega, \mathbb{R}^n)$, we have

$$\|\bar{v}(\,\cdot\,, u(\,\cdot\,))\|^q \mathcal{A}(Du(\,\cdot\,)) + \|\bar{\Phi}(\,\cdot\,, u(\,\cdot\,))\|^q \mathcal{A}(Du(\,\cdot\,)) \in L^1(\Omega).$$

Now we use the trivial estimate $\|\bar{v}(x, y)\| \leqslant \|g(x, y)\|$ for all $x \in \Omega, y \in \mathbb{R}^n$, and area formula (6.4) to conclude

$$\|\bar{v}\mathcal{A}^{\frac{1}{q}}(Du)\|_q^q = \int_\Omega \|\bar{v}(x, u(x))\|^q \mathcal{A}(Du(x)) \, \mathrm{d}x$$
$$\leqslant \int_\Omega \|g(x, u(x))\|^q \mathcal{A}(Du(x)) \, \mathrm{d}x = \int_{\mathcal{G}_u} \|g\|^q(z) \mathrm{d}\mathcal{H}^m(z) = \|g\|_{L^q(\mathcal{G}_u)}^q.$$

With analogous estimates, we also find

$$\|\bar{\Phi}\mathcal{A}^{\frac{1}{q}}(Du)\|_q^q \leqslant \|g\|_{L^q(\mathcal{G}_u)}^q.$$

Therefore, (6.9) holds with constant $C' = 2C$. Now assume (6.9). Choose the following sequence $g_k \in C_c^1(\Omega \times \mathbb{R}^n)$:

$$g_k(x, y) := G(x)\chi_k(y),$$

where $G \in C_c^1(\Omega, \mathbb{R}^{n+m})$, and $\chi_k \in C_c^\infty(\mathbb{R}^n)$ with $0 \leqslant \chi_k(y) \leqslant 1, \forall\, y \in \mathbb{R}^n$, $\chi_k \equiv 1$ on $B_k(0)$, $\chi_k \equiv 0$ on $B_{2k}^c(0)$ and $\|D\chi_k(y)\| \leqslant \frac{1}{k}$, for all $y \in \mathbb{R}^n$. Using again area formula (6.4), we write

$$\|g_k\|_{L^q(\mathcal{G}_u)}^q = \int_\Omega \|g_k(x, u(x))\|^q \mathcal{A}(Du(x)) \, \mathrm{d}x.$$

Monotone convergence theorem implies

$$\lim_k \|g_k\|_{L^q(\mathcal{G}_u)}^q = \|G\mathcal{A}^{\frac{1}{q}}(Du)\|_q^q.$$

Now we want to use (7.5). Using the same notation as in the statement of Lemma 7.3, i.e. splitting G into $G_1 = (G^1, \ldots, G^m)$ and $G_2 = (G^{n+1}, \ldots, G^{n+m})$, we have

$$\int_\Omega \langle B(Du(x)), D((g_k)_1(x, u(x))) \rangle \, dx$$

$$= \int_\Omega \langle B(Du(x)), D(\chi_k(u(x))G_1(x)) \rangle \, dx$$

$$= \int_\Omega \chi_k(u(x)) \langle B(Du(x)), DG_1(x) \rangle \, dx$$

$$+ \int_\Omega \langle B(Du(x)), G_1(x) \otimes (D\chi_k(u(x))Du(x)) \rangle \, dx$$

By Lemma 7.1 and the regularity of G_1, we have that

$$\|DG_1\| \|B(Du)\| \in L^1(\Omega) \quad \text{and} \quad \|G_1\| \|B(Du)\| \|Du\| \in L^1(\Omega). \qquad (7.18)$$

Since

$$\chi_k(u(x)) \langle B(Du(x)), DG_1(x) \rangle \longrightarrow \langle B(Du(x)), DG_1(x) \rangle$$

pointwise a.e. as $k \to \infty$, (7.18) tells us that we can apply dominated convergence theorem to infer

$$\lim_{k \to \infty} \int_\Omega \chi_k(u(x)) \langle B(Du(x)), DG_1(x) \rangle \, dx = \int_\Omega \langle B(Du(x)), DG_1(x) \rangle.$$

Moreover using the pointwise bound $\|D\chi_k(u(x))\| \leqslant \frac{1}{k}$,

$$\left| \int_\Omega \langle B(Du(x)), G_1(x) \otimes (D\chi_k(x)Du(x)) \rangle \, dx \right|$$

$$\leqslant \frac{1}{k} \int_\Omega \|B(Du(x))\| \|G_1(x)\| \|Du(x)\| \, dx.$$

Again through (7.18), we infer that the last term converges to 0. This implies that

$$\int_\Omega \langle B(Du(x)), D((g_k)_1(x, u(x))) \rangle \, dx \longrightarrow \int_\Omega \langle B(Du(x)), DG_1(x) \rangle \, dx$$

$$\text{as } k \longrightarrow \infty.$$

In a completely analogous way,

$$\int_\Omega \langle A(Du(x)), D((g_k)_2(x, u(x))) \rangle \, dx \longrightarrow \int_\Omega \langle A(Du(x)), DG_2(x) \rangle \, dx \, dx$$

$$\text{as } k \longrightarrow \infty.$$

Now (7.5) and the previous computations yield

$$\int_\Omega \langle A(Du(x)), DG_2(x)\rangle \, \mathrm{d}x + \int_\Omega \langle B(Du(x)), DG_1(x)\rangle \, \mathrm{d}x$$

$$= \lim_{k\to\infty} \left[\int_\Omega \langle A(Du(x)), D(g_k)_2(x)\rangle \, \mathrm{d}x + \int_\Omega \langle B(Du(x)), D(g_k)_1(x)\rangle \right] \mathrm{d}x$$

$$= \lim_{k\to\infty} \delta_\Psi(\llbracket \mathcal{G}_u \rrbracket)(g_k) \leqslant C' \lim_k \|g_k\|_{L^q(\mathcal{G}_u)} = C' \|G\mathcal{A}^{\frac{1}{q}}(Du)\|_q,$$

and it is immediate to see that this implies (6.8) with constant $\overline{C}' = C'$.

8. Some open questions

We list here a series of questions related to the topic of the present paper. Firstly, as already explained in the introduction, the main question which motivated the investigations of this paper is the following widely open question.

QUESTION 8.1. — *Is it possible to prove an analog of W. Allard's celebrated regularity theorem* [1] *if we consider strongly elliptic integrands (in the sense of Almgren)* Ψ *on Grassmanian?*

The answer to this question is far from being immediate. A major obstacle is the lack of the monotonicity formula, [2]. Actually most of the proof in [1] can be carried over if one know the validity of a Michael–Simon inequality. More precisely, consider a rectifiable varifold $V = \theta \llbracket \Gamma \rrbracket$ with density bounded below (e.g. $\theta \geqslant 1$) and anisotropic variation $\delta_\Psi V$ which is bounded in $L^1(\theta \mathcal{H}^m \llcorner \Gamma)$, i.e.

$$\delta_\Psi(g) = \int_\Gamma H_\Psi \cdot g \, \theta \, \mathrm{d}\mathcal{H}^m$$

for some $H_\Psi \in L^1$. The anisotropic Michael–Simon inequality would then take the conjectural form

$$\left(\int_\Gamma h^{\frac{m}{m-1}} \theta \, \mathrm{d}\mathcal{H}^m \right)^{\frac{m-1}{m}} \leqslant C \int_\Gamma \theta |\nabla^\Gamma h| + C \int_\Gamma |h| \, |H_\Psi| \, \theta \, \mathrm{d}\mathcal{H}^m \quad \forall \, h \in C_c^1. \ (8.1)$$

QUESTION 8.2. — *Is it possible to prove a Michael–Simon inequality as* (8.1) *for (at least some) anisotropic energies?*

Of course, Question 8.1 has its counterpart on graphs, which amounts to extend the partial regularity of Evans for minimizers to stationary points.

QUESTION 8.3. — *Is it possible to extend the partial regularity theorem of* [12] *to Lipschitz graphs that are stationary with respect to strongly polyconvex (or quasiconvex) energies?*

Answering these questions in this generality seems out of reach at the moment. It is however possible to formulate several interesting intermediate questions, many of which are related to the "differential inclusions point of view" adopted in the present paper.

First of all we could consider stronger assumptions on the integrand Ψ. In the recent paper [9], A. De Rosa, the second named author and F. Ghiraldin introduce the so-called Atomic Condition. Such condition characterizes those energies for which (the appropriate extension of) Allard's rectifiability result holds. The following question is thus natural (see the forthcoming paper [11] for results in this direction):

QUESTION 8.4. — *What is the counterpart of the Atomic Condition for functionals on graphs and what can be concluded from it in the graphical case?*

Secondly, a possible approach to Question 8.1 is a continuation-type argument on the space of all energies. Since the area functional has a particular status, the following question is particularly relevant.

QUESTION 8.5. — *Does an Allard type result holds for integrands which are sufficiently close to the area?*

In the forthcoming paper, [30], the fourth named author proves a partial result in the above direction. Using methods coming from the theory of differential inclusion, [30] shows that graphs with small Lipschitz constant that are stationary with respect to functions sufficiently close to the area are regular. These results, other than the one in the present paper, seem to point to partial regularity for stationary varifolds (or graphs), as opposed to the situation of [23, 26].

We note that a key step in the proof of Evans' partial regularity theorem is the so called Caccioppoli inequality which, roughly, reads as follows: for a minimizer u defined on B_2

$$\int_{B_1} |Du - Da|^2 \leqslant C \int_{B_2} |u - a|^2$$

for all affine functions $a(x) = b + Ax$. The geometric counterpart of this estimate is used by Almgren in its partial regularity theorem for currents minimizing anisotropic energies, [5]. These inequalities are obtained by direct comparison with suitable competitors. A similar estimate is obtained, by purely PDE techniques, by Allard in the case of stationary varifolds and it is one of the key step in establishing his regularity theorem for stationary varifolds. For co-dimension one stationary varifolds which are stationary with respect to anisotropic *convex* integrands, a similar inequality is known

to hold true, [3]. However, in general co-dimension, no condition on the integrand it is known to ensure its validity, not even in neighbourhoods of the area integrand, $\Psi = 1$.

QUESTION 8.6. — *Which conditions on the integrands f or Ψ ensures the validity of a Cacciopoli type inequality for stationary points?*

In [13], it is proved that for differential inclusions of the form

$$Du \in K,$$

where K is a compact set of $\mathbb{R}^{2\times 2}$ that does not contain T_4 configurations, compactness properties hold. In particular, if $\sup_j \|u_j\|_{W^{1,p}(B_1(0))} < +\infty$ for some $p > 1$, then there exists a subsequence u_{j_k} such that u_{j_k} converges strongly in $W^{1,q}(B_1(0))$ for every $q < p$. This kind of compactness property can actually be used to prove partial regularity of solutions to elliptic systems of PDEs. The strategy of [13] does not apply directly to the higher dimensional case and motivates the following question

QUESTION 8.7. — *Let $f \in C^1(\mathbb{R}^{n\times m})$ be a strictly polyconvex function and $K_f \subset \mathbb{R}^{(2n+m)\times m}$. Suppose $W_j : \Omega \to \mathbb{R}^{(2n+m)\times m}$ is a sequence of maps such that $\sup_j \|W_j\|_\infty < +\infty$, W_j satisfies the linear system (2.2), and*

$$\mathrm{dist}(W_j(x), K_f) \rightharpoonup 0$$

in the weak topology of L^p. Then, up to subsequences, does W_j converge strongly in $W^{1,p}$?

To formulate the next questions, let us recall the following definitions, see, for instance, [19], or [22, Section 4.4]. A function

$$F : \mathbb{R}^{n\times m} \longrightarrow \mathbb{R}$$

is said to be rank-one convex if $h(t) := F(X + tY)$ is a convex function, for every $X, Y \in \mathbb{R}^{n\times m}$ and $\mathrm{rank}(Y) = 1$. It is said to be quasiconvex if $\forall\, \Phi \in C_c^\infty(B_1(0), \mathbb{R}^n), B_1(0) \subset \mathbb{R}^m$ and $X \in \mathbb{R}^{n\times m}$, one has

$$\fint_{B_1(0)} F(X + D\Phi(x))\,\mathrm{d}x \geqslant F(X).$$

For compact sets $K \subset \mathbb{R}^{n\times m}$, we define

$$K^{rc} = \left\{ X \in K : \begin{array}{l} F(X) \leqslant 0, \forall\, F : \mathbb{R}^{n\times m} \longrightarrow \mathbb{R} \text{ rank-one convex s.t.} \\ \hphantom{F(X) \leqslant 0, \forall\, F :} F(Z) \leqslant 0, \forall\, Z \in K \end{array} \right\}$$

and analogously K^{qc} (resp. K^{pc}) where one uses quasiconvex (resp. polyconvex) functions instead of rank-one convex functions. Moveover, one has the following chain of inclusions

$$K \subseteq K^{rc} \subseteq K^{qc} \subseteq K^{pc}. \tag{8.2}$$

A necessary condition for compactness to hold is that $K^{qc} = K$ so that in particular $K^{rc} = K$. These results are consequence of the theory of Young Measures and in particular of the abstract result of [18]. For a thorough explanation, we refer once again the reader to [22, Section 4].

As discussed at the beginning of Section 2.3, in dimension 2 the stationarity of a graph is equivalent to solve

$$DW \in K_f := \left\{ A \in \mathbb{R}^{(2n+2)\times 2} : A = \begin{pmatrix} X \\ Df(X)J \\ X^T Df(X)J - f(X)J \end{pmatrix} \right\}, \quad (8.3)$$

for some $W \in W^{1,\infty}(\Omega, \mathbb{R}^{2n+2})$, where J is the symplectic matrix

$$\begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}.$$

Therefore we can ask

QUESTION 8.8. — *Let $f \in C^1(\mathbb{R}^{n\times 2})$ be a strictly polyconvex function. Is it true that $(K_f \cap \overline{B}_R(0))^{rc} = K_f \cap \overline{B}_R(0)$ for every $R > 0$?*

The same question can be generalized to $m > 2$ using the wave cone Λ_{dc} of Definition 2.4. In analogy with rank-one convex functions, we can introduce Λ_{dc}-convex functions

DEFINITION 8.9. — *A function $F : \mathbb{R}^{(2n+m)\times m} \to \mathbb{R}$ is Λ_{dc}-convex if $h(t) := F(X + tY)$ is a convex function for every $X \in \mathbb{R}^{(2n+m)\times m}$ and $Y \in \Lambda_{dc}$. We also define, for a compact set $K \subset \mathbb{R}^{(2n+m)\times m}$ the Λ_{dc}-convex hull*

$$K^{dc} := \left\{ X \in K : \begin{array}{l} F(X) \leqslant 0, \forall \, F : \mathbb{R}^{(2n+m\times m)} \Lambda_{dc}\text{-convex s.t.} \\ \qquad F(Z) \leqslant 0, \forall \, Z \in K \end{array} \right\}.$$

The multi-dimensional analogue of Question 8.8 is then the following:

QUESTION 8.10. — *Let $f \in C^1(\mathbb{R}^{n\times m})$ be a strictly polyconvex function and $R > 0$. Does*

$$(K_f \cap \overline{B}_R(0))^{dc} = K_f \cap \overline{B}_R(0)?$$

Appendix A. Proof of Proposition 6.4

First, by [16, Section 1.5,Theorem 1], one has that if $w \in W^{1,m}(\Omega, \mathbb{R}^{m+n})$, then for every measurable set $A \subset \Omega$ and every measurable function $g : \mathbb{R}^{m+n} \to \mathbb{R}$ for which

$$g(w(\,\cdot\,))J_w(\,\cdot\,) \in L^1(A), \quad (A.1)$$

it holds

$$\int_A g(w(x)) J_w(x) \, \mathrm{d}x = \int_{\mathbb{R}^{m+n}} g(z) N(w, A, z) \, \mathrm{d}\mathcal{H}^m(z),$$

where

$$J_w(x) = \sqrt{\det(Dw(x)^T Dw(x))}$$

$$\text{and} \quad N(w, A, z) := \#\{x : x \in A \cap A_D(w), w(x) = z\}.$$

We want to apply this result with

$$A = \mathcal{L}_u, \quad w(x) = v(x), \quad g(x, y) := F(v(x), T_{v(x)} \mathcal{G}_u), \quad \forall \, x \in \Omega, y \in \mathbb{R}^m.$$

In this way, it is straightforward by the fact that $\mathcal{R}_u = \mathcal{L}_u \cap A_D(u)$ and the definition of $v(x)$ that $N(v, \mathcal{L}_u, z) = 1$ for $\mathcal{H}^m \llcorner \mathcal{G}_u$ and $N(v, \mathcal{L}_u, z) = 0$ if $z \notin \mathcal{G}_u$. Hence:

$$\int_{\mathbb{R}^{m+n}} g(z) N(w, A, z) \, \mathrm{d}\mathcal{H}^m(z) = \int_{\mathcal{G}_u} F(v(x), T_{v(x)} \mathcal{G}_u) \, \mathrm{d}\mathcal{H}^m(z).$$

Moreover, since $|\Omega \setminus \mathcal{L}_u| = 0$ and $J_w(x) = \mathcal{A}(Du(x))$, we also find

$$\int_{\mathcal{L}_u} g(w(x)) J_w(x) \, \mathrm{d}x = \int_\Omega F(v(x), T_{v(x)} \mathcal{G}_u) \mathcal{A}(Du(x)) \, \mathrm{d}x.$$

Since $u \in W^{1,m}(\Omega, \mathbb{R}^n)$ and $F \in C_b(\Omega \times \mathbb{R}^n \times \mathbb{G}_0)$, (A.1) is fulfilled and we can apply the aforementioned result to obtain the desired equality (6.4).

Appendix B. Proof of Lemma 7.1

First of all we compute $D\mathcal{A}(X)$. Recall the notation on multi-indices introduced in Definition 3.5 and the definition of the matrix $\overline{\mathrm{cof}(X^Z)^T}$ in the proof of Proposition 4.1. Then, since

$$\mathcal{A}(X) = \sqrt{1 + \|X\|^2 + \sum_{2 \leqslant r \leqslant \min\{m,n\}} \sum_{Z \in \mathcal{A}_r} \det(X_Z)^2},$$

we have

$$D\mathcal{A}(X) = \frac{X + \sum_{2 \leqslant r \leqslant \min\{m,n\}} \sum_{Z \in \mathcal{A}_r} \det(X_Z) \overline{\mathrm{cof}(X^Z)^T}}{\mathcal{A}(X)},$$

$$\forall \, X \in \mathbb{R}^{n \times m}. \quad \text{(B.1)}$$

Next, we observe that by the chain rule

$$D(\Psi(h(X)))_{ij} = \sum_{1 \leqslant \alpha, \beta \leqslant m+n} (\partial_{\alpha\beta} \Psi)(h(X)) \partial_{ij} h_{\alpha\beta}(X),$$

hence
$$D(\Psi(h(X))) = \sum_{1 \leqslant \alpha,\beta \leqslant m+n} (\partial_{\alpha\beta}\Psi)(h(X))Dh_{\alpha\beta}(X). \qquad \text{(B.2)}$$

We can therefore write
$$A(X) = \Psi(h(X))D\mathcal{A}(X) + \mathcal{A}(X)D(\Psi(h(X)))$$
$$= \Psi(h(X))D\mathcal{A}(X) + \mathcal{A}(X) \sum_{1 \leqslant \alpha,\beta \leqslant m+n} (\partial_{\alpha\beta}\Psi)(h(X))Dh_{\alpha\beta}(X)$$

and
$$B(X) = \Psi(h(X))(-X^T D\mathcal{A}(X) + \mathcal{A}(X)\,\mathrm{id}_m)$$
$$+ \mathcal{A}(X) \sum_{1 \leqslant \alpha,\beta \leqslant m+n} (\partial_{\alpha\beta}\Psi)(h(X))X^T Dh_{\alpha\beta}(X).$$

Since $\mathbb{G}(m, m+n)$ is compact, we have that both $\Psi(h(X))$ and $(D\Psi)(h(X))$ are bounded in $L^\infty(\mathbb{R}^{n\times m})$ by a constant $c > 0$ and using (B.1), we can bound
$$\Psi(h(X))\|D\mathcal{A}(X)\| \lesssim \|X\|^{\min\{m,n\}-1}.$$

Moreover, for every $X \in \mathbb{R}^{n\times m}$, $2 \leqslant r \leqslant \min\{m,n\}$ and $Z \in \mathcal{A}_r$, we have
$$X^T \overline{\mathrm{cof}(X^Z)^T} = \det(X_Z)I_Z,$$

where, if δ_{ab} denotes Kronecker's delta and Z has the form $Z = (i_1, \ldots, i_r, j_1, \ldots, j_r)$, I_Z is the $m \times m$ matrix defined as
$$(I_Z)_{ij} = \begin{cases} 0, & \text{if } i \neq i_a \text{ or } j \neq j_b, \forall\, a, b, \\ \delta_{ab}, & \text{if } i = i_a, j = j_b. \end{cases}$$

Therefore
$$- X^T D\mathcal{A}(X) + \mathcal{A}(X)\,\mathrm{id}_m$$
$$= -\frac{X^T X + \sum_{2 \leqslant r \leqslant \min\{m,n\}} \sum_{Z \in \mathcal{A}_r} \det^2(X_Z)I_Z - \mathcal{A}^2(X)\,\mathrm{id}_m}{\mathcal{A}(X)}. \qquad \text{(B.3)}$$

If $n \leqslant m-1$, then the best way to estimate the previous expression is simply
$$\|X^T D\mathcal{A}(X) - \mathcal{A}(X)\,\mathrm{id}_m\| \lesssim 1 + \|X\|^n.$$

On the other hand, if $n \geqslant m$, then for $Z \in \mathcal{A}_m$ we have $I_Z = \mathrm{id}_m$, hence (B.3) becomes
$$- X^T D\mathcal{A}(X) + \mathcal{A}(X)\,\mathrm{id}_m$$
$$= -\frac{X^T X + \sum_{2 \leqslant r \leqslant \min\{m,n\}} \sum_{Z \in \mathcal{A}_r} \det^2(X_Z)I_Z - \mathcal{A}^2(X)\,\mathrm{id}_m}{\mathcal{A}(X)}$$
$$= -\frac{X^T X - (1 + \|X\|^2)\,\mathrm{id}_m + \sum_{2 \leqslant r \leqslant m-1} \sum_{Z \in \mathcal{A}_r} \det^2(X_Z)(I_Z - \mathrm{id}_m)}{\mathcal{A}(X)}.$$

In this case
$$\|X^T D\mathcal{A}(X) - \mathcal{A}(X)\operatorname{id}_m\| \lesssim 1 + \|X\|^{m-1}.$$
To conclude the proof of the Lemma, we still need to prove that for every $1 \leqslant i, j \leqslant m+n$,
$$\begin{aligned}
\mathcal{A}(X)\|Dh_{ij}(X)\| &\lesssim 1 + \|X\|^{\min\{n,m\}-1}, \\
\mathcal{A}(X)\|X^T Dh_{ij}(X)\| &\lesssim 1 + \|X\|^{\min\{m-1,n\}}.
\end{aligned} \tag{B.4}$$
To perform the computation, we need to divide it into cases corresponding to the four blocks of the matrix $h(X)$ as written in (6.1). To this end, recall the notation
$$S(X) = (\operatorname{id}_m + X^T X)^{-1},$$
and moreover notice that $h(X)$ is symmetric, therefore we just need to prove (B.4) in the case $i \leqslant j$. Another useful fact is the following. First notice that for every matrices $N \in O(n)$, $M \in O(m)$ ($O(k)$ is the group of orthogonal matrices of order k), one has
$$S(NXM) = M^T S(X) M.$$
From an easy computation we then conclude that, for every $1 \leqslant i, j \leqslant m+n$ and for every $X \in \mathbb{R}^{n \times m}$, $N \in O(n)$, $M \in O(m)$,
$$\|Dh_{ij}(X)\| \lesssim \sum_{1 \leqslant a,b \leqslant m+n} \|Dh_{ab}(NXM)\| \tag{B.5}$$
$$\|X^T Dh_{ij}(X)\| \lesssim \sum_{1 \leqslant a,b \leqslant m+n} \|(NXM)^T Dh_{ab}(NXM)\|. \tag{B.6}$$
Since also $\mathcal{A}(X) = \mathcal{A}(NXM)$, $\forall\, X \in \mathbb{R}^{n \times m}, M \in O(m), N \in O(n)$, (B.5) and (B.6) tell us that we can check estimates (B.4) just on matrices $Y := NXM$ with two additional hypotheses. Fix $X \in \mathbb{R}^{n \times m}$, define $Z = XM$ and denote the j-th column of a matrix $A \in \mathbb{R}^{n \times m}$ with A^j. First, by a suitable choice of M, we can make sure that $Y^T Y = Z^T Z = M^T X^T X M$ is diagonal. Once this choice is made, if $n \geqslant m$, then we choose $N = \operatorname{id}_n$. Otherwise, if $n < m$, then we observe that at most n of the columns of Z are non-zero, let these be $Z^{j_1}, \dots, Z^{j_n}$ and let us define $J := \{j_1, \dots, j_n\}$ with $1 \leqslant j_1 < j_2 < \cdots < j_n \leqslant m$. If for some j_k we have $Z^{j_k} = 0$, then we set $N = \operatorname{id}_n$. Otherwise, the $n \times n$ matrix V formed using $Z^{j_1}, \dots, Z^{j_n}$ has columns that are pairwise orthogonal and nonzero, hence there exists $O \in O(n)$ such that
$$V = OD,$$
with D diagonal. In this case, we choose $N = O^T$, so that the resulting Y has the property that
$$Y^j = \begin{cases} y_\ell e_\ell & \text{if } j = j_\ell, j_\ell \in \{j_1, \dots, j_n\}, \\ 0 & \text{otherwise,} \end{cases}$$

where $y_j \in \mathbb{R}$ and e_ℓ are the vectors of the canonical basis of $\mathbb{R}^n$. Notice that this choice of M and N also implies that

$$\mathcal{A}(Y) = \sqrt{\prod_{i=1}^{m}(1 + \|Y^i\|^2)} \text{ and } S(Y)$$

$$= \mathrm{diag}((1 + \|Y^1\|^2)^{-1}, \ldots, (1 + \|Y^m\|^2)^{-1}).$$

We call (HP) these assumptions on the matrix $Y \in \mathbb{R}^{n \times m}$.

First case, $1 \leqslant i \leqslant j \leqslant m$. — In this case, $h_{ij} = S_{ij}$. We have

$$\sum_{1 \leqslant k \leqslant m} S_{ik}^{-1} S_{kj} = \delta_{ij},$$

hence, taking a derivative,

$$\sum_{1 \leqslant k \leqslant m} \partial_{ab} S_{ik}^{-1} S_{kj} + \sum_{1 \leqslant k \leqslant m} S_{ik}^{-1} \partial_{ab} S_{jk} = 0.$$

We can invert the previous relation to get

$$\partial_{ab} S_{kl} = - \sum_{1 \leqslant c,d \leqslant m} S_{kc} S_{ld} \partial_{ab} S_{cd}^{-1}. \tag{B.7}$$

Finally, since $S_{ik}^{-1} = \delta_{ik} + \sum_{1 \leqslant l \leqslant m} x_{li} x_{lk}$, we have

$$\partial_{ab} S_{ik}^{-1} = \sum_{1 \leqslant c \leqslant m} \delta_{ci}^{ab} x_{ck} + \sum_{1 \leqslant c \leqslant m} \delta_{ck}^{ab} x_{ci},$$

where the symbol $\delta_{\alpha\beta}^{cd} = 0$ if $\alpha \neq c$ or $\beta \neq d$, otherwise $\delta_{\alpha\beta}^{cd} = \delta_{\alpha\beta}^{\alpha\beta} = 1$. We can therefore use (B.7) to write

$$\partial_{ab} S_{ij} = - \sum_{1 \leqslant k,l \leqslant m} S_{ik} S_{jl} \left(\sum_{1 \leqslant c \leqslant m} \delta_{ck}^{ab} x_{cl} + \sum_{1 \leqslant c \leqslant m} \delta_{cl}^{ab} x_{ck} \right)$$

$$= - \sum_{1 \leqslant k,l,c \leqslant m} S_{ik} S_{jl} \delta_{ck}^{ab} x_{cl} - \sum_{1 \leqslant k,l,c \leqslant m} \delta_{cl}^{ab} x_{ck} S_{ik} S_{jl} \tag{B.8}$$

$$= - \sum_{1 \leqslant l \leqslant m} \left(S_{ib} S_{jl} x_{al} + x_{al} S_{il} S_{jb} \right).$$

Moreover,

$$(X^T D S_{ij})_{cd} = \sum_{1 \leqslant a \leqslant n} x_{ac} \partial_{ad} S_{ij}$$

$$= - \sum_{1 \leqslant l \leqslant m, 1 \leqslant a \leqslant n} \left(S_{id} S_{jl} x_{al} x_{ad} + x_{ad} x_{al} S_{il} S_{jd} \right).$$

Now we use our previous observations (B.5) and (B.6) to consider Y satisfying (HP), so that in particular $Y^T Y$ is diagonal. In this case, we have

$$|\partial_{ab} S_{ij}(Y)| \leqslant \sum_{1 \leqslant l \leqslant m} \left(|S_{ib} S_{jl} y_{al}| + |y_{al} S_{il} S_{jb}| \right).$$

For every $1 \leqslant i, b, j, l \leqslant m, 1 \leqslant a \leqslant n$,

$$\mathcal{A}(Y)|S_{ib} S_{jl} y_{al}| \leqslant \sqrt{\prod_{c=1}^{m}(1 + \|Y^i\|^2)} \frac{|y_{al}|}{(1 + \|Y^b\|^2)(1 + \|Y^l\|^2)}.$$

Let us explain in detail how to get the desired estimate (B.4) in this case. Notice that either Y^l is 0, and in this case there is nothing to prove, or $Y^l \neq 0$. Thanks to (HP), in Y there are at most $\min\{m, n\}$ non-zero columns. First let $m \leqslant n$, then:

$$\sqrt{\prod_{c=1}^{m}(1 + \|Y^c\|^2)} \frac{|y_{al}|}{(1 + \|Y^b\|^2)(1 + \|Y^l\|^2)} \lesssim \sqrt{\prod_{c=1}^{m}(1 + \|Y^c\|^2)} \frac{1}{\sqrt{1 + \|Y^l\|^2}}$$
$$\lesssim 1 + \|Y\|^{m-1}.$$

If $n < m$ and J is the set on indices corresponding to non-zero columns, we are in the hypothesis in which $l \in J$. Therefore we have

$$\sqrt{\prod_{c=1}^{m}(1 + \|Y^c\|^2)} \frac{|y_{al}|}{(1 + \|Y^b\|^2)(1 + \|Y^l\|^2)} \lesssim \sqrt{\prod_{c \in J}(1 + \|Y^c\|^2)} \frac{1}{\sqrt{1 + \|Y^l\|^2}}$$
$$\lesssim 1 + \|Y\|^{m-1}.$$

This proves that

$$\|Dh_{ij}(Y)\| \lesssim 1 + \|Y\|^{\min\{m,n\}-1} \text{ for } 1 \leqslant i, j \leqslant m. \tag{B.9}$$

We also have

$$\mathcal{A}(Y)|(Y^T D S_{ij})_{cd}(Y)| \leqslant \mathcal{A}(Y) \sum_{1 \leqslant l \leqslant m, 1 \leqslant a \leqslant n} \left(|S_{id} S_{jl} y_{al} y_{ad}| + |y_{ad} y_{al} S_{il} S_{jd}| \right).$$

Analogously to the previous case, we estimate for every $1 \leqslant i, d, j, l \leqslant m, 1 \leqslant a \leqslant n$,

$$\mathcal{A}(Y)|S_{id} S_{jl} y_{al} y_{ad}| \leqslant \sqrt{\prod_{c=1}^{m}(1 + \|Y^c\|^2)} \frac{|y_{al}||y_{ad}|}{(1 + \|Y^d\|^2)(1 + \|Y^l\|^2)},$$

and the desired estimate is obtained with a reasoning completely analogous to the one of (B.9). This concludes the proof of this case.

Second case, $1 \leqslant i \leqslant m < m+1 \leqslant j \leqslant m+n.$ — From now on we use $m+j$ rather than j for the corresponding index. We thus have

$$h_{ij+m}(X) = (S(X)X^T)_{ij} = \sum_{k=1}^{m} S_{ik}x_{jk}\,.$$

We compute the derivative using (B.8):

$$\partial_{ab}h_{ij+m}(X) = \sum_{k=1}^{m} \delta_{jk}^{ab}S_{ik} + \sum_{k=1}^{m} x_{jk}\partial_{ab}S_{ik}$$

$$= \sum_{k=1}^{m} \delta_{jk}^{ab}S_{ik} - \sum_{1 \leqslant l,k \leqslant m} (S_{ib}S_{kl}x_{al}x_{jk} + x_{al}x_{jk}S_{il}S_{kb})\,,$$

and also

$$(X^T Dh_{ij+m}(X))_{ab}$$

$$= \sum_{1 \leqslant c \leqslant n} x_{ca}\partial_{cb}h_{ij}$$

$$= \sum_{1 \leqslant k \leqslant m, 1 \leqslant c \leqslant n} x_{ca}\delta_{jk}^{cb}S_{ik} - \sum_{1 \leqslant l,k \leqslant m, 1 \leqslant c \leqslant n} (x_{ca}S_{ib}S_{kl}x_{cl}x_{jk} + x_{ca}x_{cl}x_{jk}S_{il}S_{kb})$$

$$= x_{ja}S_{ib} - \sum_{1 \leqslant l,k \leqslant m, 1 \leqslant c \leqslant n} (x_{ca}S_{ib}S_{kl}x_{cl}x_{jk} + x_{ca}x_{cl}x_{jk}S_{il}S_{kb})$$

Since $S^{-1}(X) = \mathrm{id}_m + X^T X$,

$$\delta_{ij} = \sum_{1 \leqslant k \leqslant m} S_{ik}(\delta_{kj} + \sum_{1 \leqslant c \leqslant n} x_{ck}x_{cj}) = S_{ij} + \sum_{1 \leqslant k \leqslant m, 1 \leqslant c \leqslant n} S_{ik}x_{ck}x_{cj}, \quad \text{(B.10)}$$

hence we can rewrite

$$(X^T Dh_{ij+m}(X))_{ab}$$

$$= x_{ja}S_{ib} - \sum_{1 \leqslant l,k \leqslant m, 1 \leqslant c \leqslant n} (x_{ca}S_{ib}S_{kl}x_{cl}x_{jk} + x_{ca}x_{cl}x_{jk}S_{il}S_{kb})$$

$$= x_{ja}S_{ib} - \sum_{k=1}^{m} \left(S_{ib}x_{jk} \sum_{1 \leqslant l \leqslant m, 1 \leqslant c \leqslant n} x_{ca}S_{kl}x_{cl} + x_{jk}S_{kb} \sum_{1 \leqslant l \leqslant m, 1 \leqslant c \leqslant n} x_{ca}x_{cl}S_{il} \right)$$

$$= x_{ja}S_{ib} - \sum_{k=1}^{m} S_{ib}x_{jk}(\delta_{ka} - S_{ka}) - \sum_{k=1}^{m} x_{jk}S_{kb}(\delta_{ai} - S_{ai})$$

$$= \sum_{k=1}^{m} S_{ib}x_{jk}S_{ka} + \sum_{k=1}^{m} x_{jk}S_{kb}\delta_{ai} + \sum_{k=1}^{m} x_{jk}S_{kb}S_{ai}. \quad \text{(B.11)}$$

Now we evaluate the previous expressions at Y satisfying (HP). Using the fact that $Y^T Y$ is diagonal, we simplify:

$$\partial_{ab} h_{ij+m}(Y) = \sum_{k=1}^{m} \delta_{jk}^{ab} S_{ik} - \sum_{1 \leqslant l, k \leqslant m} (S_{ib} S_{kl} y_{al} y_{jk} + y_{al} y_{jk} S_{il} S_{kb})$$

$$= \delta_{ja} \delta_{ib} S_{ii} - \sum_{1 \leqslant k \leqslant m} \delta_{ib} S_{ii} S_{kk} y_{ak} y_{jk} - y_{ai} y_{jb} S_{ii} S_{bb} \tag{B.12}$$

First, let $m \leqslant n$. Then, using that for every $1 \leqslant a, j, k \leqslant n$ we have

$$|S_{kk} y_{ak} y_{jk}| \leqslant 1,$$

we estimate

$$\mathcal{A}(Y) |\partial_{ab} h_{ij+m}(Y)|$$

$$\leqslant \mathcal{A}(Y) |S_{ii}| + \mathcal{A}(Y) \sum_{1 \leqslant k \leqslant m} |S_{ii} S_{kk} y_{ak} y_{jk}| + \mathcal{A}(Y) y_{ai} y_{jb} S_{ii} S_{bb}$$

$$\lesssim \frac{\mathcal{A}(Y)}{1 + \|Y^i\|^2} + \frac{\mathcal{A}(Y)}{\sqrt{1 + \|Y^i\|^2} \sqrt{1 + \|Y^b\|^2}} \lesssim 1 + \|Y\|^{m-1}.$$

If $n < m$, let $J = \{j_1, \ldots, j_n\}$ be the set of indices defined in (HP). If there exists ℓ such that $Z^{j_\ell} = 0$, then

$$\mathcal{A}(Y) = \sqrt{\prod_{t \in J}(1 + \|Y^t\|^2)} = \sqrt{\prod_{t \in J \setminus \{j_k\}}(1 + \|Y^t\|^2)} \lesssim 1 + \|Y\|^{n-1}$$

and

$$|\partial_{ab} h_{ij+m}(Y)| \leqslant S_{ii} + \sum_{1 \leqslant k \leqslant m} S_{ii} S_{kk} |y_{ak} y_{jk}| + |y_{ai} y_{jb}| S_{ii} S_{bb} \lesssim 1,$$

therefore

$$\mathcal{A}(Y) |\partial_{ab} h_{ij}(Y)| \lesssim 1 + \|Y\|^{n-1}.$$

Hence we are just left with the case $n < m$ and $Y^{j_\ell} \neq 0$ for every $1 \leqslant \ell \leqslant n$. If this is the case, (HP) implies that $y_{kj_\ell} = \delta_{k\ell} y_{\ell j_\ell}$, for $1 \leqslant k \leqslant n$, and $y_{kj} = 0$ if $j \notin J$ and $1 \leqslant k \leqslant n$. Therefore, recalling (B.12),

$$\partial_{ab} h_{ij+m}(Y) = \begin{cases} S_{ii} - S_{ii} S_{j_a j_a} y_{a j_a}^2 - y_{ai} y_{jb} S_{ii} S_{bb} & \text{if } j = a, i = b, \\ -y_{ai} y_{jb} S_{ii} S_{bb} & \text{otherwise.} \end{cases}$$

In the first case, if $j = a, i = b$, we have

$$S_{j_a j_a} = \frac{1}{1 + \|Y^{j_a}\|^2} = \frac{1}{1 + y_{a j_a}^2},$$

hence

$$1 - S_{j_a j_a} y_{a j_a}^2 = 1 - \frac{y_{a j_a}^2}{1 + y_{a j_a}^2} = \frac{1}{1 + y_{a j_a}^2} = \frac{1}{1 + \|Y^{j_a}\|^2}$$

that implies

$$\partial_{ab}h_{ij+m}(Y) = \frac{1}{1 + \|Y^{j_a}\|^2} - \frac{y_{ai}y_{jb}}{(1 + \|Y^i\|^2)(1 + \|Y^b\|^2)},$$

and it is now easy to see that

$$\mathcal{A}(Y)|\partial_{ab}h_{ij+m}(Y)| \lesssim 1 + \|Y\|^{n-1}.$$

Since if $j \neq a$ or $b \neq i$, $\partial_{ab}h_{ij+m}(Y) = -y_{ai}y_{jb}S_{ii}S_{bb}$, the same estimate follows. To finish the second case, we still need to show that

$$\mathcal{A}(Y)|(Y^T Dh_{ij+m}(Y))_{ab}| \lesssim 1 + \|Y\|^{\min\{m,n\}-1}.$$

To do so, we recall (B.11) to estimate

$$|(Y^T Dh_{ij+m}(Y))_{ab}| \leqslant \sum_{k=1}^{m} S_{ib}|y_{jk}|S_{ka} + \sum_{k=1}^{m} |y_{jk}|S_{kb}\delta_{ai} + \sum_{k=1}^{m} |y_{jk}|S_{kb}S_{ai}.$$

With similar computations to the one to prove (B.9), we estimate for $1 \leqslant i, b, a, k \leqslant m, 1 \leqslant j \leqslant n$,

$$\mathcal{A}(Y)S_{ib}|y_{jk}|S_{ka} \leqslant \begin{cases} 0 & \text{if } Y^k = 0 \text{ or } k \neq a, \\ \sqrt{\prod_{l \neq k}(1 + \|Y^l\|^2)} & \text{otherwise,} \end{cases}$$

that implies

$$\mathcal{A}(Y)S_{ib}|y_{jk}|S_{ka} \lesssim 1 + \|Y\|^{\min\{m,n\}-1}.$$

Finally, since also for every $1 \leqslant j \leqslant n$, $1 \leqslant k, b \leqslant m$

$$\mathcal{A}(Y)|y_{jk}|S_{kb} \leqslant \begin{cases} 0 & \text{if } Y^k = 0 \text{ or } k \neq b, \\ \sqrt{\prod_{l \neq k}(1 + \|Y^l\|^2)} & \text{otherwise,} \end{cases}$$

we find

$$\mathcal{A}(Y)|y_{jk}|S_{kb} \lesssim 1 + \|Y\|^{\min\{m,n\}-1}, \qquad \forall\, 1 \leqslant k, b \leqslant m,\ 1 \leqslant j \leqslant n.$$

This completes the proof of the second case.

Third case, $m+1 \leqslant i \leqslant j \leqslant m+n$. — As above we use $m+i$ and $m+j$ in place of i and j. The indices i and j will then satisfy $1 \leqslant i \leqslant j \leqslant n$ and we have

$$h_{i+m,j+m}(X) = (XS(X)X^T)_{ij} = \sum_{1 \leqslant l, k \leqslant m} x_{il}S_{lk}x_{jk}.$$

We compute the derivative using (B.8):

$$
\begin{aligned}
\partial_{ab} & h_{i+m,j+m}(X) \\
&= \sum_{1 \leqslant l,k \leqslant m} \delta_{il}^{ab} S_{lk} x_{jk} + \sum_{1 \leqslant l,k \leqslant m} \delta_{jk}^{ab} S_{lk} x_{il} + \sum_{1 \leqslant l,k \leqslant m} x_{il} \partial_{ab} S_{lk} x_{jk} \\
&= \sum_{1 \leqslant k \leqslant m} \delta_{ia} S_{bk} x_{jk} + \sum_{1 \leqslant l \leqslant m} \delta_{ja} S_{lb} x_{il} \\
&\quad - \sum_{1 \leqslant l,k,c \leqslant m} S_{lb} S_{kc} x_{ac} x_{il} x_{jk} - \sum_{1 \leqslant l,k,c \leqslant m} x_{ac} S_{lc} S_{kb} x_{il} x_{jk}.
\end{aligned}
$$

Moreover,

$$
\begin{aligned}
(X^T & Dh_{i+m,j+m}(X))_{ab} \\
&= \sum_{1 \leqslant d \leqslant n} x_{da} \partial_{db} h_{ij} \\
&= \sum_{1 \leqslant d \leqslant n, 1 \leqslant k \leqslant m} \delta_{id} x_{da} S_{bk} x_{jk} + \sum_{1 \leqslant d \leqslant n, 1 \leqslant l \leqslant m} \delta_{jd} S_{lb} x_{il} x_{da} \qquad \text{(B.13)} \\
&\quad - \sum_{1 \leqslant c,l,k \leqslant m, 1 \leqslant d \leqslant n} S_{lb} S_{kc} x_{dc} x_{il} x_{jk} x_{da} \\
&\quad - \sum_{1 \leqslant c,l,k \leqslant m, 1 \leqslant d \leqslant n} x_{da} x_{dc} S_{lc} S_{kb} x_{il} x_{jk}.
\end{aligned}
$$

By (B.10), we have, for every $1 \leqslant i,j \leqslant m$

$$
\sum_{1 \leqslant k \leqslant m, 1 \leqslant d \leqslant n} S_{ik} x_{dk} x_{dj} = \delta_{ij} - S_{ij}.
$$

Hence we can rewrite in (B.13):

$$
\begin{aligned}
(X^T & Dh_{i+m,j+m}(X))_{ab} \\
&= \sum_{1 \leqslant d \leqslant n} x_{da} \partial_{db} h_{ij} \\
&= \sum_{1 \leqslant d \leqslant n, 1 \leqslant k \leqslant m} \delta_{id} x_{da} S_{bk} x_{jk} + \sum_{1 \leqslant d \leqslant n, 1 \leqslant l \leqslant m} \delta_{jd} S_{lb} x_{il} x_{da} \\
&\quad - \sum_{1 \leqslant l,k \leqslant m} S_{lb} x_{il} x_{jk} (\delta_{ka} - S_{ka}) - \sum_{1 \leqslant l,k \leqslant n} S_{kb} x_{il} x_{jk} (\delta_{la} - S_{la})
\end{aligned}
$$

$$= \sum_{1\leqslant d\leqslant n,1\leqslant k\leqslant m} \delta_{id}x_{da}S_{bk}x_{jk} + \sum_{1\leqslant d\leqslant n,1\leqslant l\leqslant m} \delta_{jd}S_{lb}x_{il}x_{da}$$

$$- \sum_{1\leqslant l,k\leqslant m} S_{lb}x_{il}x_{jk}\delta_{ka} - \sum_{1\leqslant l,k\leqslant m} S_{kb}x_{il}x_{jk}\delta_{la}$$

$$+ \sum_{1\leqslant l,k\leqslant m} S_{lb}x_{il}x_{jk}S_{ka} + \sum_{1\leqslant l,k\leqslant m} S_{kb}x_{il}x_{jk}S_{la}$$

$$= \sum_{1\leqslant k\leqslant m} x_{ia}S_{bk}x_{jk} + \sum_{1\leqslant l\leqslant m} S_{lb}x_{il}x_{ja}$$

$$- \sum_{1\leqslant l\leqslant m} S_{lb}x_{il}x_{ja} - \sum_{1\leqslant k\leqslant m} S_{kb}x_{ia}x_{jk}$$

$$+ \sum_{1\leqslant l,k\leqslant m} S_{lb}x_{il}x_{jk}S_{ka} + \sum_{1\leqslant l,k\leqslant m} S_{kb}x_{il}x_{jk}S_{la}$$

$$= \sum_{1\leqslant l,k\leqslant m} S_{lb}x_{il}x_{jk}S_{ka} + \sum_{1\leqslant l,k\leqslant m} S_{kb}x_{il}x_{jk}S_{la}.$$

Consider once again Y fulfilling (HP). Then:

$$\partial_{ab}h_{i+m,j+m}(Y) = \sum_{1\leqslant k\leqslant m} \delta_{ia}S_{bk}y_{jk} + \sum_{1\leqslant l\leqslant m} \delta_{ja}S_{lb}y_{il}$$

$$- \sum_{1\leqslant l,k,c\leqslant m} S_{lb}S_{kc}y_{ac}y_{il}y_{jk} - \sum_{1\leqslant l,k,c\leqslant m} x_{ac}S_{lc}S_{kb}y_{il}y_{jk}.$$

Since Y^TY is diagonal, this expression simplifies as:

$$\partial_{ab}h_{i+m,j+m}(Y) = \delta_{ia}S_{bb}y_{jb} + \delta_{ja}S_{bb}y_{ib}$$

$$- \sum_{1\leqslant c\leqslant m} S_{bb}S_{cc}y_{ac}y_{ib}y_{jc} - \sum_{1\leqslant c\leqslant m} y_{ac}S_{cc}S_{bb}y_{ic}y_{jb}.$$

For every $1\leqslant b\leqslant m, 1\leqslant j\leqslant n$,

$$\mathcal{A}(Y)S_{bb}|y_{jb}| \leqslant \begin{cases} 0 & \text{if } Y^b = 0, \\ \frac{\mathcal{A}(Y)}{\sqrt{1+\|Y^b\|^2}} & \text{otherwise.} \end{cases} \tag{B.14}$$

This yields

$$\mathcal{A}(Y)S_{bb}|y_{jb}| \lesssim 1 + \|Y\|^{\min\{m,n\}-1}.$$

To prove that

$$\mathcal{A}(Y)|\partial_{ab}h_{i+mj+m}(Y)| \lesssim 1 + \|Y\|^{\min\{m,n\}-1}, \tag{B.15}$$

we still need to estimate terms of the form

$$\mathcal{A}(Y)S_{bb}S_{cc}|y_{ac}y_{ib}y_{jc}|,$$

for $1\leqslant b,c\leqslant m, 1\leqslant a,i,j\leqslant n$. Anyway, observe that

$$S_{cc}|y_{ac}||y_{jc}| \leqslant 1, \qquad \forall\, 1\leqslant c\leqslant m,\ 1\leqslant a,j\leqslant n,$$

– 958 –

hence

$$\mathcal{A}(Y)S_{bb}S_{cc}|y_{ac}y_{ib}y_{jc}| \leqslant \mathcal{A}(Y)S_{bb}|y_{ib}|$$

and we can therefore apply again estimate (B.14) to deduce (B.15). To finish the proof of this case and of the present Lemma, we still need to show that

$$|(Y^T Dh_{i+m,j+m}(Y))_{ab}| \lesssim 1 + \|Y\|^{\min\{m,n\}-1}. \tag{B.16}$$

To do so, recall that

$$(Y^T Dh_{i+m,j+m}(Y))_{ab} = \sum_{1 \leqslant l,k \leqslant m} S_{lb}y_{il}y_{jk}S_{ka} + \sum_{1 \leqslant l,k \leqslant m} S_{kb}y_{il}y_{jk}S_{la}$$

$$= S_{bb}S_{aa}y_{ib}y_{ja} + S_{bb}S_{aa}y_{ia}y_{jb},$$

but now, for every $1 \leqslant a,b \leqslant m, 1 \leqslant i,j \leqslant n$,

$$\mathcal{A}(Y)S_{bb}S_{aa}|y_{ib}y_{ja}| \leqslant \begin{cases} 0 & \text{if } Y^b = 0 \text{ or } Y^a = 0, \\ \dfrac{\mathcal{A}(Y)}{\sqrt{1+\|Y^b\|^2}\sqrt{1+\|Y^a\|^2}} & \text{otherwise.} \end{cases}$$

The proof of (B.16) is now analogous to the one of (B.15).

Bibliography

[1] W. K. ALLARD, "On the first variation of a varifold", *Ann. Math.* **95** (1972), p. 417-491.

[2] ———, "A characterization of the Area Integrand", in *Convegni del Gennaio e del Maggio del 1973*, Symposia Mathematica, vol. XIV, Academic Press Inc., 1974, p. 429-444.

[3] ———, "An integrality theorem and a regularity theorem for surfaces whose first variation with respect to a parametric elliptic integrand is controlled", in *Geometric Measure Theory and the Calculus of Variations*, Proceedings of Symposia in Pure Mathematics, vol. 44, American Mathematical Society, 1986, p. 1-28.

[4] F. J. ALMGREN, JR, "The homotopy groups of the integral cycle groups", *Topology* **1** (1962), no. 4, p. 257-299.

[5] ———, "Existence and regularity almost everywhere of solutions to elliptic variational problems among surfaces of varying topological type and singularity structure", *Ann. Math.* **87** (1968), p. 321-391.

[6] L. AMBROSIO, N. FUSCO & D. PALLARA, *Functions of Bounded Variation and Free Discontinuity Problems*, Oxford Mathematical Monographs, Clarendon Press, 2000.

[7] C. DE LELLIS, "Allard's interior regularity theorem: an invitation to stationary Varifolds", 2017, http://cvgmt.sns.it/paper/3454/.

[8] C. DE LELLIS & L. SZÉKELYHIDI, "The Euler equations as a differential inclusion", *Ann. Math.* **170** (2009), no. 3, p. 1417-1436.

[9] G. DE PHILIPPIS, A. DE ROSA & F. GHIRALDIN, "Rectifiability of Varifolds with Locally Bounded First Variation with Respect to Anisotropic Surface Energies", *Commun. Pure Appl. Math.* **71** (2017), no. 6, p. 1123-1148.

[10] A. DE ROSA & S. KOLASIŃSKI, "Equivalence of the ellipticity conditions for geometric variational problems", *Commun. Pure Appl. Anal.* **73** (2020), no. 11, p. 2473-2515.

[11] A. DE ROSA & R. TIONE, "On the atomic condition and compactness", in preparation, 2019.

[12] L. C. Evans, "Quasiconvexity and Partial Regularity in the Calculus of Variations", *Arch. Ration. Mech. Anal.* **95** (1986), no. 3, p. 227-252.

[13] D. Faraco & L. Székelyhidi, "Tartar's conjecture and localization of the quasiconvex hull in $\mathbb{R}^{2 \times 2}$", *Acta Math.* **200** (2008), p. 279-305.

[14] H. Federer, *Geometric measure theory*, Grundlehren der Mathematischen Wissenschaften, vol. 153, Springer, 1969.

[15] C. Förster & L. Székelyhidi, "T_5-Configurations and non-rigid sets of matrices", *Calc. Var. Partial Differ. Equ.* **57** (2018), no. 1, article no. 19 (12 pages).

[16] M. Giaquinta, G. Modica & J. Soucek, *Cartesian currents in the calculus of variations I. Cartesian currents*, Ergebnisse der Mathematik und ihrer Grenzgebiete. 3. Folge., vol. 37, Springer, 1998.

[17] ———, *Cartesian currents in the calculus of variations II. Variational integrals*, Ergebnisse der Mathematik und ihrer Grenzgebiete. 3. Folge., vol. 38, Springer, 1998.

[18] D. Kinderlehrer & P. Pedregal, "Characterizations of young measures generated by gradients", *Arch. Ration. Mech. Anal.* **115** (1991), no. 4, p. 329-365.

[19] B. Kirchheim, S. Müller & V. Šverák, "Studying Nonlinear pde by Geometry in Matrix Space", in *Geometric analysis and nonlinear partial differential equations*, Springer, 2003, p. 347-395.

[20] J. Kristensen & A. Taheri, "Partial Regularity of Strong Local Minimizers in the Multi-Dimensional Calculus of Variations", *Arch. Ration. Mech. Anal.* **170** (2003), no. 1, p. 63-89.

[21] H. B. Lawson, Jr & R. Osserman, "Non-existence, non-uniqueness and irregularity of solutions to the minimal surface system", *Acta Math.* **139** (1977), p. 1-17.

[22] S. Müller, "Variational models for microstructure and phase transitions", in *Calculus of variations and geometric evolution problems*, Lecture Notes in Mathematics, vol. 1713, Springer, 1999, p. 85-210.

[23] S. Müller & V. Šverák, "Convex integration for Lipschitz mappings and counterexamples to regularity", *Ann. Math.* **157** (2003), no. 3, p. 715-742.

[24] V. Scheffer, "Regularity and irregularity of solutions to nonlinear second order elliptic systems of partial differential equations and inequalities", PhD Thesis, Princeton University, USA, 1974.

[25] L. Simon, *Lectures on Geometric Measure Theory*, Australian National University, 2008.

[26] L. Székelyhidi, "The Regularity of Critical Points of Polyconvex Functionals", *Arch. Ration. Mech. Anal.* **172** (2004), no. 1, p. 133-152.

[27] ———, "Rank-one convex hulls in $\mathbb{R}^{2 \times 2}$", *Calc. Var. Partial Differ. Equ.* **28** (2007), no. 4, p. 545-546.

[28] ———, "Private communication", 2018.

[29] A. Taheri, "Quasiconvexity and uniqueness of stationary points in the multi-dimensional calculus of variations", *Proc. Am. Math. Soc.* **131** (2003), no. 10, p. 3101-3107.

[30] R. Tione, "Minimal graphs and differential inclusions", `https://arxiv.org/abs/2002.02157`, 2020.

[31] L. C. Young, "Surfaces paramétriques généralisées", *Bull. Soc. Math. Fr.* **79** (1951), p. 59-84.