

Article

A Pareto–Pontryagin Maximum Principle for Optimal Control

Alberto Lovison ^{1,*}  and Franco Cardin ² 
¹ Dipartimento di Matematica, Politecnico di Milano, Piazza Leonardo da Vinci, 32, 20133 Milano, Italy

² Dipartimento di Matematica “Tullio Levi-Civita”, Università di Padova, Via Trieste, 63, 35121 Padova, Italy; cardin@math.unipd.it

* Correspondence: alberto.lovison@polimi.it

Abstract: In this paper, an attempt to unify two important lines of thought in applied optimization is proposed. We wish to integrate the well-known (dynamic) theory of Pontryagin optimal control with the Pareto optimization (of the static type), involving the maximization/minimization of a non-trivial number of functions or functionals, Pontryagin optimal control offers the definitive theoretical device for the dynamic realization of the objectives to be optimized. The Pareto theory is undoubtedly less known in mathematical literature, even if it was studied in topological and variational details (Morse theory) by Stephen Smale. This reunification, obviously partial, presents new conceptual problems; therefore, a basic review is necessary and desirable. After this review, we define and unify the two theories. Finally, we propose a Pontryagin extension of a recent multiobjective optimization application to the evolution of trees and the related anatomy of the *xylems*. This work is intended as the first contribution to a series to be developed by the authors on this subject.

Keywords: calculus of variations; optimal control; multiobjective optimization; Pareto optimality



Citation: Lovison, A.; Cardin, F. A Pareto–Pontryagin Maximum Principle for Optimal Control.

Symmetry **2022**, *14*, 1169.

<https://doi.org/10.3390/sym14061169>

sym14061169

Academic Editor: Jan Awrejcewicz

Received: 5 May 2022

Accepted: 31 May 2022

Published: 6 June 2022

Publisher’s Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction and Motivations

We start by giving some motivation for the Pareto–Pontryagin investigation that we are proposing. An interesting and simple example for understanding the notion of multiobjective control appears to be the growth process of a tree. Like almost every living being, a tree has to accomplish different tasks, i.e., pursue different objectives, to survive and thrive. Different conflicting objectives for optimizing a tree may be the maximization of water transported from the roots to the leaves and the minimization of the resources (carbon) used to build the vessels, called xylems, and the different tree structures. The tree exerts a physiological control on the growth process through hormonal emission, under the physical constraints imposed by the environment, i.e., the available resources such as lighting, carbon dioxide, humidity, and nutrients. We see that this problem, described in detail in Sections 4.2 and 5.4, can be modeled as an optimal control problem, where different objectives are involved and need to be optimized, maximized, or minimized in a suitable compromised way, by the so-called *Pareto optimization*.

When at least two conflicting objectives are involved, the growth process can produce a wide range (possibly an infinity) of different optimal “balanced” solutions, named Pareto optimal solutions. If we compare two of them, we will observe that one may be better at transporting water, as in broad-leaved plants, while the other could result in a more parsimonious use of carbon for vessel construction, as in conifers. Such diversification in the solutions reflects the different relative importance between objectives observed in different environments or ecological conditions and consists of a strong thrust toward biodiversity. We believe that such a dynamic complexity and evolution can be framed in a Pareto–Pontryagin theory, where the term *Pontryagin* proposes controlled dynamics, while the term *Pareto* denotes vector optimization, realizing definitively a so-called multiobjective control problem. In Section 5 this problem will be precisely formulated and discussed in detail.

A similar situation is observed in engineering design. Establishing a project to develop a vehicle, or a space mission, usually entails maximizing travel speed and minimizing fuel consumption at the same time. A multipurpose vehicle should be able to perform those different tasks in many different “mixing proportions” representing different operational situations, where the travel speed and the fuel consumption may vary in their relative importance, although not completely disappearing in favour of the remaining objective.

In addition to the above application examples providing motivation to this area of research, we need to recall that the genuine origin of multiobjective optimization (MO) dates back to the work in mathematical economy by F. Y. Edgeworth [1], extended and popularized by the engineer and economist Vilfredo Pareto [2,3]. The optimal solutions of MO problems are named after him.

The same concepts and philosophy of multiobjective optimization may be applied to the calculus of variations and to the mathematical control theory, although explicit development of the theory and applications is not yet systematic even though it has grown sensibly in recent years [4–9].

In the largest part of the present literature on multiobjective optimal control (MOC), the control problem is reduced to a finite dimensional multiobjective optimization problem. This is performed in the following three steps. First, the problem is translated to a multiobjective calculus of variations problem by a control-to-state operator (applicable when the differential constraint has a unique solution for every choice of the controls). Second, the time-dependent problem is discretized, obtaining a large but finite dimensional multiobjective optimization problem that can finally be tackled by standard numerical methods. In some limited cases (e.g., [8,10]), specific concepts and results of optimal control are investigated in their possible adaptation to the multiple objectives case.

The contributions of the present work are the following. In Section 2, we survey some important elements of mathematical control theory, with a special emphasis on the geometric aspects. We stress in particular the use of a singular Legendre transformation. Further investigations and analysis on the role of Lagrange multipliers and on symplectic aspects of the theory are proposed in the Appendices A and B. In Section 3, we introduce multiobjective optimization using the optimization of a function under constraints as the motivating framework. We discuss the extension to calculus of variations in Section 4, along with some applications. Finally, in Section 5, we discuss the basic necessary conditions on optimal solutions for multiobjective control problems, as they should extend the classical results contained in the Pontryagin theory [11,12]. In particular, we investigate the multiobjective extensions of the Pontryagin maximum principle, both in the Lagrange form and in the Hamilton(–Pontryagin) form. To obtain such results, we rely on the existence of non-negative Lagrange multipliers, as proven in the following first-order proposition contained in a fundamental work by Stephen Smale:

Proposition 1 (*‘First-order proposition’, Smale [13,14]*). *A necessary and sufficient condition for x to belong to set of Pareto optima related to k functions (or functionals) $\mathcal{J}_\alpha$ is that one of the following equivalent conditions hold:*

- (a) *The set of vectors $d\mathcal{J}_\alpha(x)$, $\alpha = 1, \dots, k$, do not all lie in the same open half-space [in the dual linear space $T^*\mathbb{R}^k$];*
- (b) *There exist $\lambda_\alpha \geq 0$, $\alpha = 1, \dots, k$, not all zero, such that $\sum_\alpha \lambda_\alpha d\mathcal{J}_\alpha(x) = 0$.*

This result essentially is originally contained in the multiobjective version of the Karush–Kuhn–Tucker theorem [15,16]. The importance of Smale’s investigations is the generalization of the Morse theory to multiple objectives, arising in his dynamical systems and global analysis setting. Notably, [9,17–19] investigated in more detail the fallout of these seminal works of Smale, extending the classification of Pareto critical points by Morse-type indices and suggesting extensions to multiple objectives of existing results in critical point theory. The existence of such Lagrange multipliers is crucial for bridging the multiobjective theory to the standard scalar optimization theory and thereby obtaining the

desired corresponding results. This is obtained by writing a unique scalar function with the linear convex combination given by the multipliers of the family of objective functions or functionals.

2. A Simple Kind of Geometrical Portrait of the Optimization

A Singular Legendre Transformation

Coming back to the origins is always a good thing to do. Often, when we try to revisit ancestral aspects of modern theories, we are able to reach some more deep meanings of the matter we are concerned with. A good example is rebuilding the Pontryagin Hamiltonian function in an optimal control environment by following some classical lines of thought. We recall, especially with treating the *Mayer problem*, see, e.g., [20,21], that when the differential constraint is $\dot{x} = \varphi(x, u)$, equipped with the *variation equation* $\dot{v} = \langle \nabla \varphi, v \rangle$, the standard way to produce the momenta p is to introduce the *adjoint equation*, $\langle p, v \rangle = \text{const}$, arriving to $\dot{p} = -\langle \nabla \varphi^T, p \rangle$. In such a case, the Hamiltonian function arising is $\mathcal{H} = \langle p, \varphi(x, u) \rangle$. From an earlier point of view, closer to the ideas of classical mechanics, as, e.g., in [12] and especially in [11], when dealing with the *Lagrangian problem*, the standard optimal control setting could be rewritten in the following line. We recall it briefly, for the moment without specifying the functional domains.

Optimal Control Problem:

Determine $(x(\cdot), u(\cdot))$ such that $\int_0^T L(x(t), u(t)) dt$ is minimized, under the differential constraint $\dot{x} = \varphi(x, u)$, $x(0) = x_0$. (In [11] p. 320, the symbol f_0 is used in place of L).

Basic arguments (Lagrangian multipliers theory, see Appendix A) lead us to consider equivalently the *augmented Lagrangian function*

$$\mathcal{L}(x, \lambda, \dot{x}, \dot{\lambda}; u) = L(x, u) + \lambda \cdot (\dot{x} - \varphi(x, u)) \quad \left(= \mathcal{L}(x, \lambda, \dot{x}, \underbrace{\dot{\lambda}}_{\text{absent}}; u) \right) \quad (1)$$

with Lagrange equations

$$\frac{d}{dt} \frac{\partial \mathcal{L}}{\partial \dot{x}} - \frac{\partial \mathcal{L}}{\partial x} = 0, \quad \frac{d}{dt} \frac{\partial \mathcal{L}}{\partial \dot{\lambda}} - \frac{\partial \mathcal{L}}{\partial \lambda} = 0, \quad (2)$$

that read as

$$\dot{\lambda} = \frac{\partial L}{\partial x} - \lambda \cdot \frac{\partial \varphi}{\partial x}, \quad \dot{x} = \varphi(x, u). \quad (3)$$

We stress that $(x, \lambda; \dot{x}, \dot{\lambda})$ can be naturally considered inside $T(\mathbb{R}_x^n \times \mathbb{R}_\lambda^n)$, even though $\dot{\lambda}$ are not involving $\mathcal{L}$. The Lagrangian setting $\mathcal{L}(q, \lambda, \dot{q}, \dot{\lambda}) = L(q, \dot{q}) + \lambda \cdot \varphi(q)$ is quite natural and is used for instance in [22,23]. The presence of $\dot{\lambda}$ is trivial, but the Euler–Lagrange equations related to $\mathcal{L}(q, \lambda, \dot{q}, \dot{\lambda})$ offer the exact set of dynamic equations of the new constrained system. The Legendre transformation from $T(\mathbb{R}_x^n \times \mathbb{R}_\lambda^n)$ into $T^*(\mathbb{R}_x^n \times \mathbb{R}_\lambda^n)$ is trivially definitively singular. Moreover, we proceed by crossing our fingers, recalling also that the present construction at this level forgets the auxiliary control variables u :

$$p_x = \frac{\partial \mathcal{L}}{\partial \dot{x}}(x, \lambda, \dot{x}, \dot{\lambda}; u) = \lambda, \quad p_\lambda = \frac{\partial \mathcal{L}}{\partial \dot{\lambda}}(x, \lambda, \dot{x}, \dot{\lambda}; u) = 0, \quad (4)$$

and the related Hamiltonian function reads

$$\mathcal{H}(x, \lambda, p_x, p_\lambda; u) = \dot{x} \cdot \frac{\partial \mathcal{L}}{\partial \dot{x}} + \dot{\lambda} \cdot \frac{\partial \mathcal{L}}{\partial \dot{\lambda}} - \mathcal{L}, \quad (5)$$

where we are forced to insert in this definition Equation (3)₂,

$$\mathcal{H}(x, \lambda, p_x, p_\lambda; u) = \varphi(x, u) \cdot \frac{\partial \mathcal{L}}{\partial \dot{x}} - \mathcal{L} = \varphi(x, u) \cdot \lambda - L(x, u) + \lambda \cdot (\dot{x} - \varphi(x, u)), \quad (6)$$

$$\mathcal{H}(x, \lambda, p_x, p_\lambda; u) = -L(x, u) + \varphi(x, u) \cdot \lambda \quad \left(= \mathcal{H}(x, \lambda, p_x, \underbrace{p_\lambda}_{\text{absent}}; u) \right) \quad (7)$$

Formula (7)₁ is crucial, and we will write it, after (4)₁, in the following way,

$$\mathcal{H}(x, p; u) = -L(x, u) + p \cdot \varphi(x, u), \quad (8)$$

where we have relaxed p_x into p , which coincides with λ . In other words, in a somewhat lucky and wild way we have just restored the correct *Pontryagin Hamiltonian function* (Formula (7) in [11], p. 320). This brief tale detects a suitable framework for our *Pareto–Pontryagin problem* below to foresee the correct and useful way to tackle this new problem.

Moreover, we would like to recall that the use in the mechanics of Legendre degenerate transformations as above has been well studied geometrically by Tulczyjew (see Appendix B): a Lagrangian $L(x, \dot{x})$ which does not admit a regular Legendre transformation leads in a natural way—that is, in the symplectic setting of analytic mechanics—to a Hamiltonian $H(x, p; \xi)$ together with the Equation (A29), where (A29)₃ can be thought as a sort of optimization on the new ‘control’ parameters ξ arising from the Maslov–Hörmander theory.

Symbols used

- x coordinates of the configuration space $\Omega \subseteq \mathbb{R}^n$.
- p moments in the cotangent space $T\Omega$.
- λ Lagrange multipliers, in some cases may be transformed in the moments p .
- u control parameters.
- v in Appendix A represent the vectorial unknown in fields theory with constraints.
- ξ in Appendix B, are the Maslov–Hörmander control parameters in $H(x, p; \xi)$.
- ℓ in Appendix B, are the free parameters in the Lagrangian manifold Λ .

3. Optimizing Several Functions at the Same Time

3.1. Motivation

The scope of this section is to introduce multiobjective optimization as a natural and general mindset for managing complex decision tasks that usually are modelled as constrained optimization.

In real problems, only rarely it is possible to identify a unique criterion or objective to be maximized or minimized at the expense of all other possible criteria. Very often, to keep things simple enough to be accessible with available and reasonably demanding methods, a unique objective is chosen to be improved as much as possible, while all the remaining objectives are formulated in terms of constraints. Nevertheless, such a strategy is only a first approximation of the process that usually is put in practice. Indeed, during the decision process, it is not unusual to reconsider the assumptions and, for instance, reconsider the constraint values or transform an objective function in a constraint and vice versa. This more general point of view needs to be placed in the appropriate formal framework, where suitable strategies can be devised and implemented.

3.2. Constrained and Unconstrained Optimization—Lagrange Multipliers

Let us consider a standard (scalar) optimization problem (if not explicitly specified, we always consider minimization)

$$f : \Omega \rightarrow \mathbb{R}, \quad f^{(min)} := \min_{x \in \Omega} f(x), \quad \Omega \subseteq \mathbb{R}^n. \quad (9)$$

For f continuous and Ω compact, existence and uniqueness of a global minimum value $f^{(min)}$ is guaranteed. The minimum value is realized by one or more optimal points, counterimages of $f^{(min)}$, denoted by

$$\Omega_{(min)} := \arg \min_{x \in \Omega} f(x) = f^{-1}(f^{(min)}). \quad (10)$$

To introduce multiobjective optimization, we consider in addition a parametric equality constraint defined as follows:

$$\begin{cases} \min f(x) \\ g(x) = \bar{g}, \end{cases} \quad (11)$$

with $g: \Omega \rightarrow \mathbb{R}$, at least continuous. For every fixed value $\bar{g} \in g(\Omega)$ in the range of g , we again have existence and uniqueness of the minimum value by compactness. Indeed, $g^{-1}(\bar{g})$ is a nonempty closed subset of the compact Ω ; therefore it is still a compact set. We denote by

$$f_{\bar{g}}^{(min)} := \min_{x \in g^{-1}(\bar{g})} f(x) = \min_{x \in g^{-1}(\bar{g})} f(x). \quad (12)$$

the solution of the constrained minimization problem for the constraint g fixed to the value $\bar{g}$.

Varying the Constraint Value $\bar{g}$

As long as the value $\bar{g}$ defining the constraint varies in the range of g , $g(\Omega)$, the constrained minima $f_{\bar{g}}^{(min)}$ span a parametric subset of $\text{Im}(g \times f) \subseteq \mathbb{R}^2$. Let us consider the following smooth and convex case.

Example 1. Let $\Omega = \mathbb{S}^2 \subset \mathbb{R}^3$ and consider the projections on the first two axes x_1 and x_2 :

$$\begin{cases} f(x) = \pi_2(x) = x_2, \\ g(x) = \pi_1(x) = x_1. \end{cases} \quad (13)$$

This example is illustrated in Figure 1 and reprised in Remark 3 and in Figure 4. The range of the vector function $F := g \times f : \Omega \rightarrow \mathbb{R}^2$ is the disk $\{(x_1, x_2) | x_1^2 + x_2^2 \leq 1\}$. The range of f is $[-1, 1]$; therefore, the global unconstrained minimum of f is $f^{(min)} = -1$. The range of the constraint g is again $[-1, 1]$. Let $\bar{g} \in [-1, 1]$, as in the figure. Then the solid blue line is the image through F of $g^{-1}(\bar{g})$, and the blue dot corresponds to the constrained minimum $f_{\bar{g}}^{(min)} = \bar{f}$ of f on $\Omega_{\bar{g}}$. As $\bar{g}$ spans $g(\Omega) = [-1, 1]$, the corresponding constrained minimum value $f_{\bar{g}}^{(min)}$ spans the range $[-1, 0]$. The set of points realizing the minima $\{x \in \Omega | x = \arg \min_{x \in g^{-1}(\bar{g})} f(x)\}$ is mapped on the blue dashed curve by the vector function $F = g \times f$.

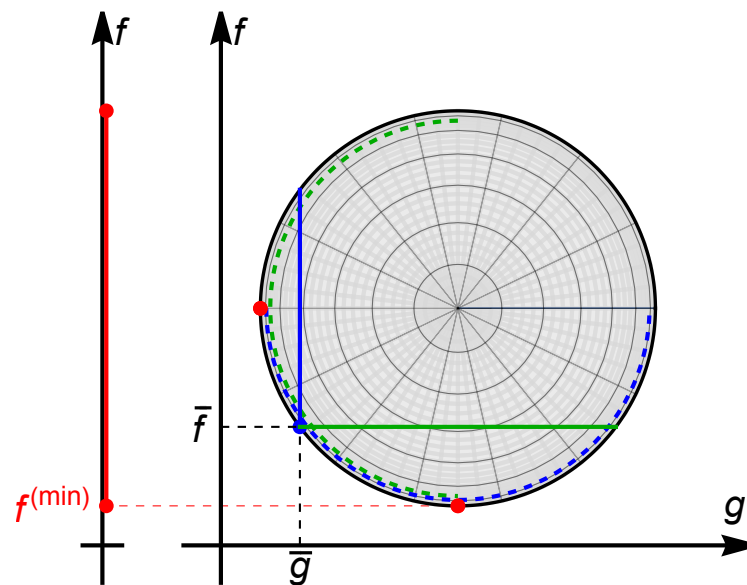


Figure 1. Left panel: The image $f(\Omega)$ with the global minimum value $f^{(min)}$ is highlighted in red. Right panel: Image set of the function $F = g \times f : \Omega \rightarrow \mathbb{R}^2$; blue solid line: image of the function f restricted to the constrained subset of Ω given by $g(x) = \bar{g}$. $\bar{f}$ is the constrained minimum; green solid line: image of the function g restricted to the constrained subset of Ω given by $f(x) = \bar{f}$; $\bar{g}$ in this case, plays the role of the constrained minimum $g_{\bar{f}}^{(min)}$; blue dashed curve: set $\mathcal{O}_1$ of constrained minima of the function f . Green dashed curve: set $\mathcal{O}_2$ of constrained minima of the function g .

3.3. Constrained Optimization and Lagrange Multipliers

When the functions are smooth and also Ω is a smooth manifold, a necessary condition for optimality under constraints is the stationarity of a suitable linear combination between the functional f and the constraint g with coefficients to be determined. Another interpretation of this fact is that the gradient of the functional f and the gradient of the constraint g have to be linearly dependent. The linear combination is called the Lagrange function:

$$\mathcal{L}_1(x, \lambda) := f(x) + \lambda g(x), \quad \lambda \in \mathbb{R}. \quad (14)$$

Theorem 1 (Standard Lagrange principle—existence of Lagrange multipliers). *If the point $\bar{x} \in \Omega$ is an extremum for the differentiable function, $f: \Omega \rightarrow \mathbb{R}$ under the constraint $g: \Omega \rightarrow \mathbb{R}$, $g(x) = \bar{g}$, then there exists a suitable number $\lambda \in \mathbb{R}$ such that*

$$\frac{d\mathcal{L}_1(\bar{x}, \lambda)}{dx} = \frac{df}{dx}(\bar{x}) + \lambda \frac{dg}{dx}(\bar{x}) = 0, \quad g(\bar{x}) = \bar{g}. \quad (15)$$

3.4. Interchanging the Roles between Functionals and Constraints

In the convex case, there exists a symmetry in the roles and the behavior of the functional and the constraint. More precisely, there exists a symmetric problem with respect to (11) that is parametrized for every possible value $\bar{f} \in f(\Omega)$:

$$\begin{cases} \min g(x), \\ f(x) = \bar{f}, \end{cases} \quad (16)$$

By writing the Lagrangian, we obtain:

$$\mathcal{L}_2(x, \mu) := g(x) + \mu f(x), \quad \mu \in \mathbb{R}, \quad (17)$$

and the Lagrange principle says that if $\bar{x}$ is a solution for (16), then:

$$\frac{d\mathcal{L}_2(\bar{x}, \mu)}{dx} = \frac{dg}{dx}(\bar{x}) + \mu \frac{df}{dx}(\bar{x}) = 0, \quad f(\bar{x}) = \bar{f}. \quad (18)$$

In relevant cases, the optimum point $\bar{x} \in \Omega$ of (11) is also the optimum point for a dual problem (16). This happens when the triple $(\bar{x}, \bar{f}, \bar{g}) \in \Omega \times f(\Omega) \times g(\Omega)$ satisfies:

$$\begin{cases} \bar{f} = \min f(x) = f(\bar{x}), \\ g(\bar{x}) = \bar{g}, \end{cases} \quad \text{if and only if} \quad \begin{cases} \bar{g} = \min g(x) = g(\bar{x}), \\ f(\bar{x}) = \bar{f}, \end{cases} \quad (19)$$

or, in other words,

$$\bar{x} = \arg \min_{x \in g^{-1}(\bar{g})} f(x) = \arg \min_{x \in f^{-1}(\bar{f})} g(x). \quad (20)$$

In a more symmetric and unified way, we can say that the point $\bar{x} \in \Omega$ is a solution of a generalized optimization problem for the two functions f and g , having as “*minimum value*” the pair $(\bar{f}, \bar{g})$. For such a generalized solution, we can write a generalized Lagrangian:

$$\mathcal{L}_3(x, \lambda_1, \lambda_2) := \lambda_1 f(x) + \lambda_2 g(x), \quad (\lambda_1, \lambda_2) \in \mathbb{R}^2, \quad (21)$$

where the Lagrange principle reads as

$$\frac{d\mathcal{L}_3(\bar{x}, \lambda_1, \lambda_2)}{dx} = \lambda_1 \frac{df}{dx}(\bar{x}) + \lambda_2 \frac{dg}{dx}(\bar{x}) = 0, \quad \text{for some } (\lambda_1, \lambda_2) \in \mathbb{R}^2. \quad (22)$$

The pair (λ_1, λ_2) can be expressed in terms of the values λ and μ related to the previous Lagrange principles (15) and (18):

$$\begin{aligned} (\lambda_1, \lambda_2) &= (1, \lambda) = \left(1, \frac{1}{\mu}\right), \quad \text{or} \quad (\lambda_1, \lambda_2) = (\mu, 1) = \left(\frac{1}{\lambda}, 1\right), \quad \text{or} \\ (\lambda_1, \lambda_2) &= \left(\frac{1}{1+\lambda}, \frac{\lambda}{1+\lambda}\right) = \left(\frac{\mu}{\mu+1}, \frac{1}{\mu+1}\right); \quad \text{therefore, } \lambda_1 + \lambda_2 = 1. \end{aligned} \quad (23)$$

The last relation appears natural because any nonzero scalar multiple of the pair (λ_1, λ_2) is equivalent in Equation (22). Therefore, a preferred choice is the pair giving a linear convex combination of f and g .

We now want to characterize the cases in which the symmetric problem (19) is solved by the same $\bar{x}$. Let us consider the product space of the ranges of both f and g , the pairs composed of a constraint value for g (resp. f) and the respective constrained minimum for the other function f (resp. g):

$$\mathcal{O}_1 := \left\{ \left(\bar{g}, \min_{x: g(x)=\bar{g}} f(x) \right) : \bar{g} \in g(\Omega) \right\} \subseteq g(\Omega) \times f(\Omega) \subseteq \mathbb{R}^2, \quad (24)$$

$$\mathcal{O}_2 := \left\{ \left(\min_{x: f(x)=\bar{f}} g(x), \bar{f} \right) : \bar{f} \in f(\Omega) \right\} \subseteq g(\Omega) \times f(\Omega) \subseteq \mathbb{R}^2. \quad (25)$$

With reference to Figure 1, $\mathcal{O}_1$ is marked with a blue dashed line, while $\mathcal{O}_2$ is marked with a green dashed line. These two subsets have a common nonempty intersection $\mathcal{O}_1 \cap \mathcal{O}_2$, which is exactly the set of pairs $(\bar{g}, \bar{f})$ for which there exists $\bar{x} \in \Omega$ such that $\bar{f} = f(\bar{x})$ and $\bar{g} = g(\bar{x})$, and that $\bar{x}$ solves both (11) and (16). We have already obtained for such a problem a symmetric unified formulation of the Lagrange principle (22). However, the symmetric optimization problem generalizing (11) and (16) has not yet been precisely defined.

We need a suitable definition of a problem of *simultaneous minimization* of both $f(x)$ and $g(x)$ or, in other words, of the vector function $F = g \times f : \Omega \rightarrow \mathbb{R}^2$, $F(x) = (g(x), f(x))$.

3.5. Multiobjective Optimization

The precise mathematical definition of a *vector* minimization problem as described in the previous paragraph owes its name to Vilfredo Pareto:

Definition 1 (Pareto optimum). A point $x \in \Omega$ is a Pareto optimum for a vector valued function $F : \Omega \rightarrow \mathbb{R}^k$, $f(x) = (f_1(x), \dots, f_k(x))$ if one of the following equivalent conditions holds:

1. There does not exist $y \in \Omega, y \neq x$, such that $f_i(y) \leq f_i(x)$ for all $i = 1, \dots, k$ and $f_j(y) < f_j(x)$ for some $j = 1, \dots, k$;
2. If there exist $y \in \Omega$ and $i = 1, \dots, k$ such that $f_i(y) < f_i(x)$, then there exist $j \neq i$ such that $f_j(x) < f_j(y)$.

The set of all Pareto optima, a subset of Ω , is called a Pareto set or a Pareto optimal set and denoted by $\mathcal{P}$. The image of the Pareto set $F(\mathcal{P}) \subseteq \mathbb{R}^k$ is called the Pareto front.

Definition 2 (Proper Pareto optimum in KT sense). A point $\hat{x} \in \Omega$ is a proper Pareto optimum for $F = f_1 \times \dots \times f_k : \Omega \rightarrow \mathbb{R}^k$ in the sense of Kuhn and Tucker [15] if there does not exist a vector $d \in T_{\hat{x}}\Omega$ such that

$$df_i(\hat{x}) \cdot d \leq 0, \quad \text{for all } i = 1, \dots, k \quad \text{and there exists } j \in \{1, \dots, k\} \quad df_j(\hat{x}) \cdot d < 0. \quad (26)$$

Remark 1. We notice that the global optimum of one of the functions $f_1, \dots, f_k$ is not automatically a Pareto optimum. This is true when there exists a unique minimizer $\hat{x}$ realizing the global minimum of, say, f_i . If the minimizer $\hat{x}$ is not unique, then it will also be a Pareto optimum depending on the value of the remaining functions on $\hat{x}$. Moreover, it is easy to check that the global minimizer of one of the functions f_i , even if unique, is not a proper Pareto optimum, at least when a nondegeneracy condition holds. Indeed, if $df_{\bar{i}}(\hat{x}) = 0$ for an index $\bar{i} \in \{1, \dots, k\}$, and $\text{corank } dF(x) = 1$ for all singular points (i.e., $\text{rank } dF(x) < \max$), it is possible to find a half space containing all the remaining gradients, and therefore to find $d \in T_{\hat{x}}\Omega$ such that $df_i(\hat{x}) \cdot d \leq 0$, for all $i = 1, \dots, k$, and $df_j(\hat{x}) \cdot d < 0$ for at least one j . For a finer characterization of critical and optimal points in multiobjective optimization see for instance [19,24].

Definition 3. A linear convex combination $\lambda_1, \dots, \lambda_k \in \mathbb{R}^{\geq 0}$, $\lambda_1 + \dots + \lambda_k = 1$, of the vector functional components $f_\lambda(x) := \lambda_1 f_1(x) + \dots + \lambda_k f_k(x)$ is called a (linear) scalarization of the vector functional f .

Proposition 2. An optimum in ordinary (scalar) sense of any scalarization $f_\lambda(x) = \lambda_1 f_1(x) + \dots + \lambda_k f_k(x)$, is a Pareto optimum for the vector functional $f = (f_1, \dots, f_k)$ if the coefficients $\lambda_1, \dots, \lambda_k$ are all nonzero (When some of the λ_i is zero then the values of the corresponding functions f_i have to be examined to detect the Pareto optima. See Remark 1).

Linear scalarizations play a critical role in multiobjective optimization. In particular, we observe that in the problem illustrated in Figure 1, the Pareto front is the intersection of the green and the blue curved lines, i.e., $f(\mathcal{P}) = \mathcal{O}_1 \cap \mathcal{O}_2$. Nevertheless, it is important to observe that linear scalarizations do not automatically produce all of the possible Pareto optima. Indeed, in the nonconvex case, there exist Pareto optima which cannot be represented as (scalar) optima of a linear scalarization.

Example 2 (Counterexample). Let us consider $P_1, P_2 \in \mathbb{R}^2$, $P_1 \neq P_2$ and the functions $f_i(x) := |x - P_i|^{\frac{1}{2}}$. Then for any $\lambda = (\lambda_1, \lambda_2) \in (\mathbb{R}^{\geq 0})^2$, $\lambda \neq 0$,

$$\arg \min_{x \in \Omega} f_\lambda(x) = \lambda_1 f_1(x) + \lambda_2 f_2(x) = P_1 \vee P_2. \quad (27)$$

Nevertheless, any $x = \lambda_1 P_1 + \lambda_2 P_2$, $\lambda_i \geq 0$, $\lambda_1 + \lambda_2 = 1$, is a Pareto optimum for $f = (f_1, f_2)$. This situation is illustrated in Figure 2.

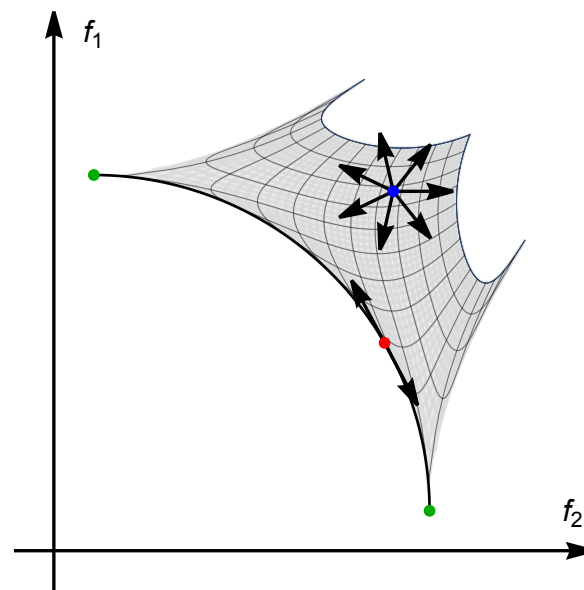


Figure 2. Image of the nonconvex vector function $F = f_2 \times f_1 : \mathbb{R}^2 \rightarrow \mathbb{R}^2$, $f_1(x) = |x - P_1|^{1/2}$, $f_2(x) = |x - P_2|^{1/2}$, with $P_1 \neq P_2$. The points in the nonconvex part of the Pareto front, the *tradeoff* optima as the red point, are Pareto optimal, but they are never scalar optima for any linear scalarization of f_1 and f_2 . Indeed, the optimum for each $\lambda_1 f_1(x) + \lambda_2 f_2(x)$ is P_1 or P_2 (green points). Nevertheless, these tradeoff optima satisfy the necessary first-order condition: there exists a linear scalarization for which they are critical points. The small arrows represent the image set of the differential dF of F , which is nonsurjective on the Pareto set. The blue point is a generic point where the differential dF is surjective, i.e., its image set is the whole $\mathbb{R}^2$.

On the other hand, every Pareto optimum is critical for a suitable linear scalarization. More precisely, we have the following fact anticipated in the introduction as Proposition 1:

Proposition 3 (Smale first-order proposition [13,14]). *If a point $\bar{x} \in \Omega$ is a (local) Pareto optimum for $f = (f_1, \dots, f_k)$, then one of the following equivalent conditions holds:*

1. *The gradients $\nabla f_1(\bar{x}), \dots, \nabla f_k(\bar{x})$ are not contained in the same open half space for any half plane in $\mathbb{R}^2$;*
2. *There exist $\lambda_1, \dots, \lambda_k \in \mathbb{R}^{\geq 0}$, not all zero, such that $\lambda_1 \nabla f_1(\bar{x}) + \dots + \lambda_k \nabla f_k(\bar{x}) = 0$.*

If one of the previous conditions hold, then $\bar{x}$ is said to be a Pareto critical point.

Interesting second-order sufficient conditions for local Pareto optimality [13] have been proven and numerically exploited for defining efficient optimization methods, e.g., [25–29]; however, they fall beyond the scope of the present work.

4. Multiobjective Calculus of Variations

Let us consider for simplicity the case of k autonomous functionals in the absence of equality or inequality constraints. Let us first consider the space of curves

$$\Gamma_{t_0, x_0}^{t_1, x_1} := \left\{ x(\cdot) \in C^1([t_0, t_1], \mathbb{R}^n) : x(t_0) = x_0, x(t_1) = x_1 \right\}.$$

Definition 4. Let $L_i : \mathbb{R}^{2n} \rightarrow \mathbb{R}$, $(x, v) \mapsto L_i(x, v)$, be at least twice the differentiable functions ($k \geq 2$), and let $\mathcal{J}_i[x(\cdot)]$ be the corresponding functionals (also called variational principles) to be minimized at the same time:

$$\mathcal{J}_i[x(\cdot)] = \int_0^T L_i(x(t), \dot{x}(t)) dt \rightarrow \inf, \quad i = 1, \dots, k, \quad x(\cdot) \in \Gamma_{t_0, x_0}^{t_1, x_1}. \quad (28)$$

We will say that a curve $\hat{x}(\cdot) \in \Gamma_{t_0, x_0}^{t_1, x_1}$ is a Pareto optimal curve or a weak Pareto extremal curve, if:

1. There does not exist another curve $x(\cdot) \in \Gamma_{t_0, x_0}^{t_1, x_1}$, such that $\mathcal{J}_i[x(\cdot)] \leq \mathcal{J}_i[\hat{x}(\cdot)]$ for all $i = 1, \dots, k$ and $\mathcal{J}_j[x(\cdot)] < \mathcal{J}_j[\hat{x}(\cdot)]$ for some $j \in \{1, \dots, k\}$;
2. For any curve $x(\cdot) \in \Gamma_{t_0, x_0}^{t_1, x_1}$, $\mathcal{J}_j[x(\cdot)] < \mathcal{J}_j[\hat{x}(\cdot)]$ for some $j \in \{1, \dots, k\}$ implies that there exist $i \neq j$ such that $\mathcal{J}_i[x(\cdot)] > \mathcal{J}_i[\hat{x}(\cdot)]$.

Choosing $\lambda_1, \dots, \lambda_k \in \mathbb{R}$, we define the combined Lagrangian:

$$L_\lambda(x(t), \dot{x}(t)) := \sum_{i=1}^k \lambda_i L_i(x(t), \dot{x}(t)), \quad (29)$$

which gives the associated combined Lagrange functional or combined variational principle:

$$\mathcal{J}_\lambda[x(\cdot)] = \int_0^T L_\lambda(x(t), \dot{x}(t)) dt. \quad (30)$$

The following Theorem is an infinite dimensional extension of the Lagrange principle.

Theorem 2 (Pareto–Lagrange for Calculus of Variations (CoV)). If $\hat{x}(\cdot) \in \Gamma_{t_0, t_1}^{x_0, x_1}$ is a Pareto optimal curve for the vector functional $\mathcal{J} = (\mathcal{J}_1, \dots, \mathcal{J}_k)$, then there exist $\lambda = (\lambda_1, \dots, \lambda_k)$, such that $\lambda_1, \dots, \lambda_k \in \mathbb{R}^\geq$, $\lambda_1 + \dots + \lambda_k = 1$, and

$$\partial_{x(\cdot)} \mathcal{J}_\lambda[\hat{x}(\cdot)] = 0, \quad \text{i.e.,} \quad -\frac{d}{dt} \frac{\partial L_\lambda}{\partial \dot{x}}(\hat{x}, \dot{\hat{x}}) + \frac{\partial L_\lambda}{\partial x}(\hat{x}, \dot{\hat{x}}) \equiv 0. \quad (31)$$

Proof. For the proof, we refer to [11] (pp. 241–245) where a single constrained functional is considered. It is necessary to reduce to a finite dimensional problem and resort to the Karush–Kuhn–Tucker Theorem. Substituting the original KKT Theorem with its multiobjective version [16] (p. 39) gives the desired result. More details will be provided in Section 5 dealing with optimal control. $\square$

Let us also consider the following “combined” Hamiltonian in the hypothesis of the convexity of $L_i(x, \dot{x})$ with respect to the $\dot{x}$ variables, for every $i = 1, \dots, k$:

$$\mathcal{H}_\lambda(x(t), p(t)) := \sup_{\dot{x} \in \mathbb{R}^n} \left[p \cdot \dot{x} - \sum_{i=1}^k \lambda_i L_i(x(t), \dot{x}(t)) \right]. \quad (32)$$

4.1. Application: Continuum Mechanics

We notice that the variational approach to the continuum mechanics of a two-phase material proposed in [30,31] can also be interpreted within the framework of multiobjective optimization. The authors consider an *Ericksen bar* [32], i.e., a one-dimensional elastic bar with a two-well nonconvex strain energy $f(u')$, with interfacial energy and an elastic foundation. The corresponding total energy functional [30] (formula (2.6), p. 1378):

$$E = \int_0^1 [f(u'(x)) + \alpha(u''(x))^2 + \beta u^2(x)] dx, \quad \alpha, \beta \in \mathbb{R}^{>0} \quad (33)$$

exhibits a large number of local equilibria. This functional is clearly a linear scalarization of three different functionals that cannot be considered separately because they give rise to a degeneracy in the solutions. Alongside a regularization of the solutions, such an approach offers an explicit framework to study the wide variety of finite-scale equilibrium microstructures observed in multi-phase solids, for instance in memory-shape alloys, in which the different material properties can be described by different values of the (small) parameters α and β .

4.2. Application: The Widened Pipe Model for Xylems

The widened pipe model (WPM) [33] of plant hydraulic evolution is a recent proposal for explaining by optimization through evolution, the origin of the profile of water vessels observed in vascular plants, called xylems. The WPM predicts that xylem conduits should be narrowest at the stem tips, widening quickly before plateauing toward the stem base (see Figure 3a).

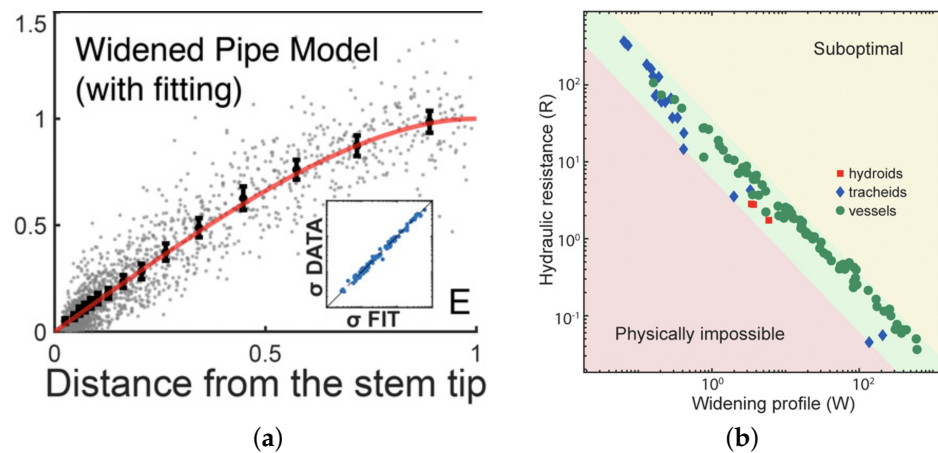


Figure 3. Panel (a): cross-sectional area of xylem conduits in a database of 103 vascular plants compared with the prediction of the widening pipe model WPM (red line); Panel (b): estimated carbon cost (W) for the xylem conduit construction versus estimated hydraulic resistance (R) for the 103 plants of the same dataset. Pictures are reproduced from [33].

With reference to the theory exposed above, the WPM consists of a multiobjective calculus of variations problem. The observed xylem profile is represented in terms of the cross-sectional area σ widening as a function of the distance from the stem of the conduit toward the roots h . The xylem profile $\sigma(h)$ is the result of a trade-off between two competing factors occurring in natural selection: one favoring rapid widening of conduits tip to base, minimizing hydraulic resistance $R[\sigma(\cdot)]$ and another favoring slow widening of conduits, minimizing carbon cost and embolism risk $W[\sigma(\cdot)]$ (see Figure 3b). The hydraulic resistance term R is derived from the Hagen–Poiseuille law for the laminar flow of a Newtonian fluid through a cylindrical pipe and consists of a term proportional to the inverse of the square of the cross-sectional area and to the length of the pipe infinitesimal element dh :

$$R[\sigma(\cdot)] := \int_{h_{\min}}^{h_{\max}} \frac{1}{\sigma(h)^2} dh. \quad (34)$$

The construction carbon cost term W is proportional to the total surface area of the xylem and is approximated with a term proportional to the square of $\sigma'(h) = \frac{d\sigma}{dh}(h)$:

$$W[\sigma(\cdot)] := \int_{h_{\min}}^{h_{\max}} \sigma'(h)^2 dh. \quad (35)$$

The tradeoff between these two competing functionals is obtained with a suitable linear convex combination, where the coefficients λ_1 and λ_2 play the role of Lagrange multipliers in the scalarized functionals. For every pair (λ_1, λ_2) , we consider the scalarized variational principle:

$$\mathcal{J}_{(\lambda_1, \lambda_2)}[\sigma(\cdot)] = \lambda_1 R[\sigma(\cdot)] + \lambda_2 W[\sigma(\cdot)] = \int_{h_{\min}}^{h_{\max}} \lambda_1 \frac{1}{\sigma(h)^2} + \lambda_2 \sigma'(h)^2 dh \rightarrow \min, \quad (36)$$

which admits as minimizer the following profile

$$\sigma(h) := \sigma_M \sqrt{\frac{h}{h_M} \left(2 - \frac{h}{h_M}\right)} \equiv \sigma_M F\left(\frac{h}{h_M}\right), \quad (37)$$

where $\sigma_M = \sigma(h_M)$ depends on the proportions between λ_1 and λ_2 , i.e., definitely from the relative importance between the two functionals R and W . $F(x) := \sqrt{x(2-x)}$ is a universal scaling function obtained by integrating the Euler–Lagrange equation associated with the variational principle (36). The presence of a universal scaling function F is remarkable and allows for fitting all the xylem profiles measured in the dataset of 103 vascular plants studied in [33] with respect to the same functional form by a simple linear rescaling of each set of data (see Figure 3a). We notice that the shortest plant in the dataset (*Dendroligotrichum dendroides*) was 35 cm tall, while the tallest individual (*Sequoia sempervirens*) reached over 100 m, i.e., the dataset involves more than two decades of values of h_{max} .

By fitting the data of each tree with the analytic solution (37) it is possible to estimate the corresponding hydraulic resistance R and carbon cost W . By representing these values on a Cartesian plane, we visualize a remarkable Pareto front with a narrow transversal disturbance which can be ascribed to measurements error, to fitting errors, or to higher order terms not considered in the model (see Figure 3b).

This application may appear unusual at first sight because of its inverse nature of detecting the optimization problem (multiobjective in this case) on the basis of the available solutions (the physical realizations of the xylem conduits in the dataset of the 103 plants). The crucial importance of multiobjective optimization for this kind of problems relies in the possibility of inferring the functional meaning of the objective functions involved in the evolution of plants process, i.e., to have an idea of what the direction pursued by nature was while shaping the characteristics of living beings. An inverse problem gives precious insights on the workings of nature. Similar attempts have been worked out and proposed in [34,35].

Compare this Section with Section 5.4 below, where a dynamical version of the same problem is considered.

5. Multiobjective Optimal Control

5.1. Necessary First Order Conditions for Multiobjective Optimal Control

Let us again consider for simplicity k functionals in the absence of equality or inequality constraints. We start by recalling the main concepts. We refer to [11] for the standard scalar theory and to [4], among others, for the multiobjective case.

Definition 5. Let us again consider k functionals (variational principles) depending on extra control variables $u \in \mathcal{U} \subseteq \mathbb{R}^m$ to be minimized at the same time:

$$\mathcal{J}_\alpha[x(\cdot), u(\cdot), t_0, t_1] = \int_{t_0}^{t_1} f_\alpha(t, x(t), u(t)) dt \rightarrow \inf, \quad \alpha = 1, \dots, k, \quad (38)$$

with shared differential constraints:

$$\begin{cases} \dot{x}_i = \varphi_i(t, x_j, u_L), & i = 1, \dots, n, \quad L = 1, \dots, m, \\ x_i(0) = x_i^0. \end{cases} \quad (39)$$

We will say that $\hat{\gamma} = (x(\cdot), u(\cdot), t_0, t_1)$ is a Pareto optimal process or weak Pareto extremal if it solves the differential constraints and if

1. There does not exist a process γ solving the differential constraints, such that $\mathcal{J}_i[\gamma] \leq \mathcal{J}_i[\hat{\gamma}]$ for all $i = 1, \dots, k$ and $\mathcal{J}_j[\gamma] < \mathcal{J}_j[\hat{\gamma}]$ for some $j \in \{1, \dots, k\}$;
2. For any process γ , solving the differential constraints, $\mathcal{J}_j[\gamma] < \mathcal{J}_j[\hat{\gamma}]$ for some $j \in \{1, \dots, k\}$ implies that there exist $i \neq j$ such that $\mathcal{J}_i[\gamma] > \mathcal{J}_i[\hat{\gamma}]$.

Fixing $\lambda_1, \dots, \lambda_k \in \mathbb{R}$, we define the combined Lagrange functional:

$$\mathcal{J}_\lambda[x(\cdot), u(\cdot), t_0, t_1; p(\cdot), \lambda_1, \dots, \lambda_k] = \int_{t_0}^{t_1} L_\lambda(t, x(t), \dot{x}(t), u(t), p(t)) dt \quad (40)$$

where the integrand, called combined Lagrangian, is the following function

$$L_\lambda(t, x(t), \dot{x}(t), p(t), u(t)) := \sum_{\alpha=1}^k \lambda_\alpha f_\alpha(t, x(t), u(t)) + p(t) \cdot (\dot{x} - \varphi(t, x(t), u(t))). \quad (41)$$

In order, we have the main results:

Theorem 3 (Pareto–Lagrange principle). If $\hat{\gamma} = (x(\cdot), u(\cdot), t_0, t_1)$ is a Pareto optimal process, then there exist $\lambda_1 \geq 0, \dots, \lambda_k \geq 0$ not all zero, (a $p(\cdot)$ also exists, determined by the differential constraints), such that

$$\partial_{x(\cdot)} \hat{\mathcal{J}}_\lambda = 0, \quad \text{i.e.,} \quad -\frac{d}{dt} \frac{\partial L_\lambda}{\partial \dot{x}}(t) + \frac{\partial L_\lambda}{\partial x}(t) \equiv 0, \quad (42)$$

$$\partial_{u(\cdot)} \hat{\mathcal{J}}_\lambda = 0, \quad \text{i.e.,} \quad -\frac{d}{dt} \frac{\partial L_\lambda}{\partial u}(t) \equiv 0, \quad (43)$$

Proof. The proof is adapted from [11] (pp. 320–325) considered for a single constrained functional. The original proof reduces the infinite dimensional problem to a finite dimensional problem and makes use of the Karush–Kuhn–Tucker Theorem. Substituting the original KKT Theorem with its multiobjective version [16] (p. 39) gives the desired result. $\square$

5.2. A Pareto–Pontryagin Maximum Principle

Let us fix $\lambda_1, \dots, \lambda_k$ and consider the following “combined” Pontryagin’s Hamiltonian:

$$\mathcal{H}_\lambda(t, x, p, u) := \frac{\partial L_\lambda}{\partial \dot{x}} \cdot \dot{x} - L_\lambda = p(t) \cdot \varphi(t, x, u) - \sum_{\alpha=1}^k \lambda_\alpha f_\alpha(t, x, u). \quad (44)$$

Theorem 4 (Pareto–Hamilton). In the above hypotheses, the following Hamilton equations hold:

$$\dot{x} = \frac{\partial \mathcal{H}_\lambda}{\partial p}, \quad \dot{p} = -\frac{\partial \mathcal{H}_\lambda}{\partial x}, \quad \frac{\partial \mathcal{H}_\lambda}{\partial u} = 0. \quad (45)$$

Remark 2. The strict analogy between (45), the Equation (A28) proposed in the Appendix B, and Equation (14) in [11] (p. 303) is remarkable.

Theorem 5 (Pareto–Pontryagin maximum principle (scalarized)). If $\hat{\gamma} = (\hat{x}(\cdot), \hat{p}(\cdot), \hat{u}(\cdot), t_0, t_1)$ is a Pareto optimal process, there exist $\lambda_1, \dots, \lambda_k \in \mathbb{R}^{\geq 0}$, $\lambda_1 + \dots + \lambda_k = 1$, such that:

$$\mathcal{H}_\lambda(t, \hat{x}(t), \hat{p}(t), \hat{u}(t)) \geq \mathcal{H}_\lambda(t, \hat{x}(t), \hat{p}(t), \omega), \quad \text{for all } \omega \in \mathcal{U}, t \in [t_0, t_1]. \quad (46)$$

Proof. The proof is exactly the Pontryagin proof [11], where the standard Lagrangian is substituted with a linear scalarization of the family of Lagrangians $L_\lambda = \lambda_1 L_1 + \dots + \lambda_k L_k$. $\square$

For the last result, we also consider the family of Hamiltonians associated with the individual functionals f_α :

$$\mathcal{H}_\alpha(t, x, p, u) := \frac{\partial L_\alpha}{\partial \dot{x}} \cdot \dot{x} - L_\alpha = p \cdot \varphi(t, x, u) - f_\alpha(t, x, u). \quad \alpha = 1, \dots, k. \quad (47)$$

It is therefore possible to formulate and prove a Pareto–Pontryagin Maximum Principle in a more genuine form without referring to the Lagrange multipliers (at least in the statement). For this last result, we adopt the *normality* assumption for the Pareto optimal process considered: an optimal process $\hat{\gamma}$ [36] is said to be *normal* if the set of Lagrange multipliers $\lambda_1, \dots, \lambda_k$ are all positive. In multiobjective optimization, this condition is associated with the concept of *proper* Pareto optimality defined in (2) (see also [15]).

Theorem 6 (Pareto–Pontryagin maximum principle (primal, i.e., unscaled)). *If $\hat{\gamma} = (\hat{x}(\cdot), \hat{p}(\cdot), \hat{u}(\cdot), t_0, t_1)$ is a normal Pareto optimal process, and if there exists an index $\alpha \in \{1, \dots, k\}$, a $t \in [t_0, t_1]$ and an $\omega \in \mathcal{U} \subseteq \mathbb{R}^m$ such that*

$$\mathcal{H}_\alpha(t, \hat{x}(t), \hat{p}(t), \hat{u}(t)) < \mathcal{H}_\alpha(t, \hat{x}(t), \hat{p}(t), \omega) \quad (48)$$

then there exists another index $\beta \neq \alpha$ such that

$$\mathcal{H}_\beta(t, \hat{x}(t), \hat{p}(t), \hat{u}(t)) > \mathcal{H}_\beta(t, \hat{x}(t), \hat{p}(t), \omega) \quad (49)$$

Proof. By the preceding Theorem 5, we have that

$$\mathcal{H}_\lambda(t, \hat{x}(t), \hat{p}(t), \hat{u}(t)) \geq \mathcal{H}_\lambda(t, \hat{x}(t), \hat{p}(t), \omega) \quad (50)$$

for all $t \in [t_0, t_1]$ and $\omega \in \mathcal{U}$. Nevertheless, we have:

$$\begin{aligned} \sum_{\alpha=1}^k \lambda_\alpha \mathcal{H}_\alpha(t, x, p, u) &= \sum_{\alpha=1}^k \lambda_\alpha p \varphi(t, x, u) - \sum_{\alpha=1}^k \lambda_\alpha f_\alpha(t, x, p, u) = \\ &= p \varphi(t, x, u) - \sum_{\alpha=1}^k \lambda_\alpha f_\alpha(t, x, p, u) = \mathcal{H}_\lambda(t, x, p, u), \end{aligned} \quad (51)$$

being the multipliers chosen such that $\sum_{\alpha=1}^k \lambda_\alpha = 1$. Let us write for simplicity $H_\alpha(u) := \mathcal{H}_\alpha(t, \hat{x}(t), \hat{p}(t), u)$ for fixed t and arbitrary $u \in \mathcal{U}$. Because of (50), we have

$$\sum \lambda_\alpha (H_\alpha(\hat{u}(t)) - H_\alpha(\omega)) \geq 0. \quad (52)$$

Assume there exists an index $\bar{\alpha}$, with $\lambda_{\bar{\alpha}} \neq 0$, such that

$$H_{\bar{\alpha}}(\hat{u}(t)) < H_{\bar{\alpha}}(\omega). \quad (53)$$

The inequality (52) can be rewritten as:

$$(0 <) H_{\bar{\alpha}}(\omega) - H_{\bar{\alpha}}(\hat{u}(t)) \leq \sum_{\alpha \neq \bar{\alpha}} \frac{\lambda_\alpha}{\lambda_{\bar{\alpha}}} (H_\alpha(\hat{u}(t)) - H_\alpha(\omega)). \quad (54)$$

Therefore, being all $\lambda_\alpha \geq 0$, there must exist at least an index $\beta \neq \bar{\alpha}$ such that

$$H_\beta(\hat{u}(t)) > H_\beta(\omega), \quad (55)$$

which is what was desired. $\square$

Remark 3. We illustrate in the convex case the meaning of the notion of normality for the Pareto optima. We start by considering the yet nontrivial case of one objective function f_1 and a constraint f_2 described in the previous Example 1 illustrated in Figure 4. If $\hat{x}$ is a minimum for f_1 constrained by $f_2(x) = \bar{f}_2$, then there will exist non-negative Lagrange multipliers λ_1 and λ_2 such that $\lambda_1 \nabla f_1(\hat{x}) + \lambda_2 \nabla f_2(\hat{x}) = 0$ with $f_2(\hat{x}) = \bar{f}_2$. If $\hat{x}$ is abnormal, then $\lambda_1 = 0$. This means that λ_2 must be nonvanishing. Therefore, $\nabla f_2(\hat{x}) = 0$, i.e., $\hat{x}$ is a critical point for the constraint function f_2 , and it must be one of the points P_2 or P_4 .

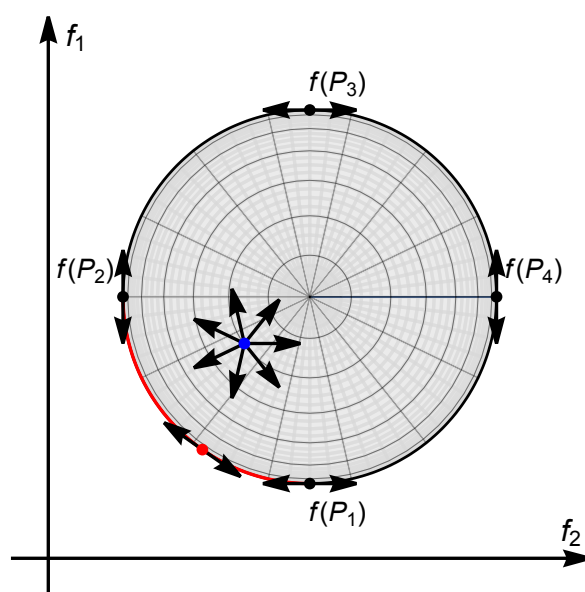


Figure 4. Image of the convex vector function $F = f_2 \times f_1$. The red curve represents the Pareto front, which is a subset of the image of the singular set of F , i.e., the set where the differential $dF : T\Omega \rightarrow T\mathbb{R}^2 \equiv \mathbb{R}^2$ is degenerate, i.e., its rank is not maximal. The small arrows going out from the highlighted points represent the image of the differential. Such an image is 1-dimensional on the points along the boundary of the image set, i.e., on the images of the singular set. Because of the first-order proposition, for Pareto optimal points there exist non-negative Lagrange multipliers giving a vanishing combination of the gradients of the objective functions $\lambda_1 \nabla f_1(\hat{x}) + \lambda_2 \nabla f_2(\hat{x}) = 0$. On the points $P_1, \dots, P_4$, one of the multipliers vanishes; therefore, the gradient relative to the other objective function is zero. Such cases are named *abnormal*. In the picture, P_2 and P_1 are abnormal Pareto optimal points. They are also the (unique and global) minimizers for the two scalar functions f_2 and f_1 , respectively, i.e., in the notation of Section 3, $f_2(P_2) = f_2^{(min)}$ and $f_1(P_1) = f_1^{(min)}$.

Now let us consider both f_1 and f_2 as objective functions and discuss the meaning of abnormal Pareto optima $\hat{x}$. The definition yields $\lambda_1 = 0$ or $\lambda_2 = 0$, i.e., the points where $\nabla f_2 = 0$ or $\nabla f_1 = 0$, respectively. Potentially, $\hat{x}$ could be one among $P_1, \dots, P_4$. Nevertheless, because we are considering Pareto optima, the choice falls only on P_1 or P_2 , i.e., the global optima of the two objective functions considered separately and without constraints. We notice that such points represent boundary points for the Pareto set.

Remark 4. It is possible to prove that the Pareto set for the functions $f_1, \dots, f_k$ is a $(k - 1)$ -dimensional stratified set in the sense of Thom. The strata of such a set are strictly related to the Pareto sets associated with subsets of $h < k$ objective functions $f_{i_1}, \dots, f_{i_h}$, which compose a sort of hierarchical $(h - 1)$ -dimensional geometrical skeleton for the full Pareto set [25–29] (In [34,35], the vertices are referred to as archetypes). In Figure 5, the generic convex case for $k = 2, 3, 4$ is illustrated, where the Pareto set is diffeomorphic to a $(k - 1)$ -simplex.

Example 3. More precisely, if we consider the case of three positive definite second-degree polynomials $f_1, f_2, f_3 : \mathbb{R}^2 \rightarrow \mathbb{R}^3$, where the respective minima P_1, P_2, P_3 are in general position, then the Pareto set is a “triangle” with curvilinear edges, where the “vertices” correspond to the minima P_1, P_2, P_3 , and the “edges” are the Pareto set for the pairs of functions $\{f_1, f_2\}, \{f_2, f_3\}, \{f_3, f_1\}$. In Figure 6, we have considered three parabolas with cylindrical symmetry:

$$f_i(x) = (x - P_i)^2, \quad P_1 = (0, 1), P_2 = \left(\frac{\sqrt{3}}{2}, -\frac{1}{2}\right), P_3 = \left(-\frac{\sqrt{3}}{2}, -\frac{1}{2}\right). \quad (56)$$

The cylindrical symmetry causes the level sets to be perfect circles centered at the points P_1, P_2, P_3 ; therefore, the Pareto set is a perfect triangle, with straight edges that in general are curved, as well as for the surface on which the Pareto set lies.

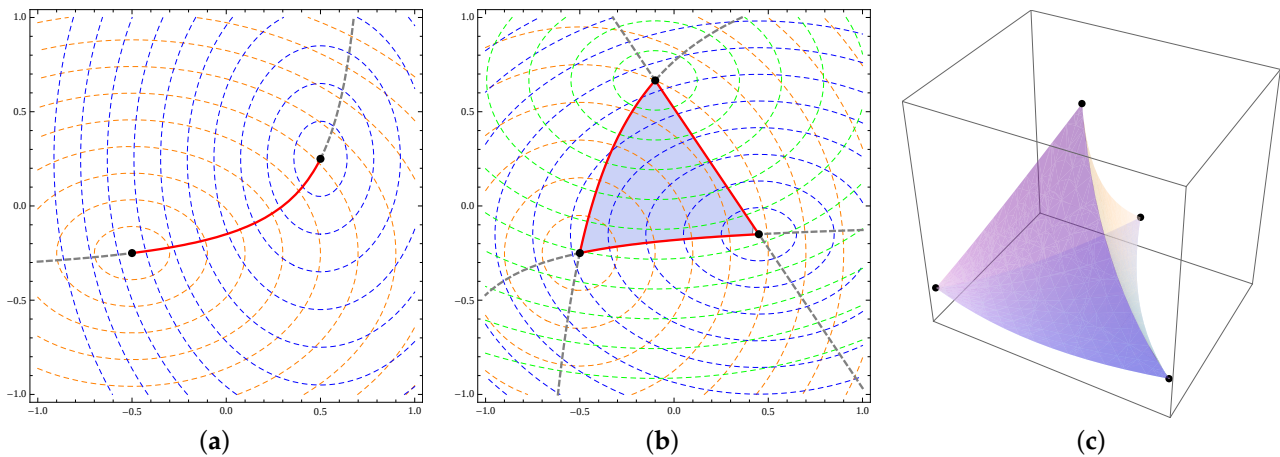


Figure 5. Pareto sets in the generic convex case: (a) two functions; (b) three functions; (c) four functions. The boundary of the Pareto set decomposes on manifolds of decreasing dimension (faces, edges, vertices, etc.) called strata, which are collections of simplexes of lower dimension. Each stratum of $h - 1$ dimension is the Pareto set for a selection of h functions among the $f_1, \dots, f_k$. The dashed lines represent the level sets of the objective functions. More on this hierarchical decomposition can be found in [28].

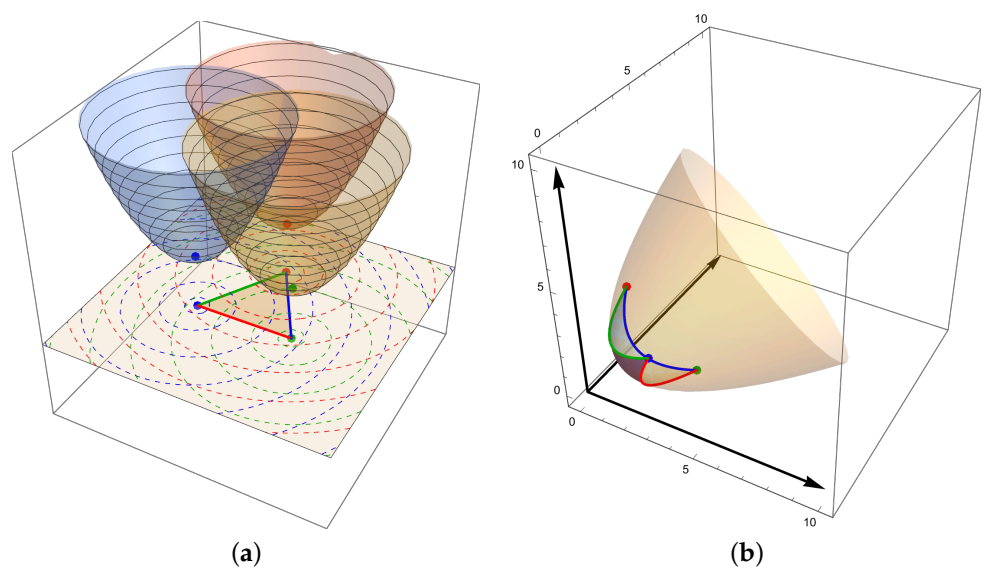


Figure 6. Pareto set and Pareto front with hierarchical stratification highlighted for the convex case with three functions. For the explicit formulation of the functions see Example 3. Panel (a): graph of the three functions, i.e., three parabolas with cylindrical symmetry with nonaligned minima. The strata of the Pareto set are highlighted in colors. The triangle is the Pareto set for the three functions together. The vertices are the optimal points for the functions considered separately (archetypes), while the edges are obtained as a Pareto set for couples of functions. Panel (b): the representation of the Pareto front, with the images of the strata highlighted with the same colors of panel (a). Different colors correspond to different functions. The highlighted and colored points correspond to the global minimal values of the functions f_1, f_2, f_3 , considered separately.

5.3. Application: The Optimal Maneuver for Racing Motorbikes

Optimizing several functionals by scalarization is a concept that has been around for quite a while, even without explicit reference to the Pareto theoretical framework. For instance, in a series of works, see [37–39] among others, the authors devised efficient numerical methods for solving the optimal control problem of obtaining the least lap time for a racing motorbike. An accurate model for a motorbike, considering also a physically realistic pilot, have to take into account several constraints. The authors transformed the constraints into penalty functions and then optimized a single functional obtained as a linear combination, with suitably tuned weights w_i , of the main physical model of the motorbike $f(t, x(t), u(t))$ and of the penalty functions $f_i(t, x(t), u(t))$:

$$I[x(\cdot), u(\cdot)] = \int_0^T \left\{ f_0(t, x(t), u(t)) + \sum_{i=1}^m w_i f_i(t, x(t), u(t)) \right\} dt, \quad (57)$$

see in particular Formula (5) in [40] (p. 117).

5.4. Application: A Growth Model for Xylematic Conduits

A Pontryagin implementation of the already mentioned Pareto xylematic conduits profile Problem 4.2 was run following the scheme below. Let us consider the following (scalarized) functional

$$F[\sigma(\cdot)] = \int_{h_{\min}}^{h_{\max}} L(\sigma(h), \sigma'(h)) dh = \int_{h_{\min}}^{h_{\max}} \left(\lambda_1 \frac{1}{\sigma^2} + \lambda_2 \frac{(\sigma')^2}{2} \right) dh \quad (58)$$

where h is the height of the xylema, σ is the area of the local circular section, and $\sigma' = d\sigma/dh$. In more detail, we can conjecture that there is a new dynamic functional, $\dot{h} = dh/dt$, $\dot{\sigma} = d\sigma/dt$,

$$\hat{F}[h(\cdot), \sigma(\cdot)] = \int_{t_{\min}}^{t_{\max}} \left(\lambda_1 \frac{1}{\sigma^2} + \lambda_2 \frac{(\dot{\sigma})^2}{2(\dot{h})^2} \right) \dot{h} dt \quad (59)$$

together with a differential constraint

$$\dot{h} = a + \varphi(h, \sigma, t; u) \quad (60)$$

As a first approximation, it is meaningful to suppose that the growing time rate of the height h of the xylems is a rather small constant, say $a > 0$, just in correspondence to $\varphi \equiv 0$. Note that in this last case, a previous pair of multipliers (λ_1, λ_2) is moved into $(a\lambda_1, \lambda_2/a)$, realizing a finer calibration of the above Pareto pair and bringing back the dynamical formulation to the previous static modellization considered in [33] and briefly recovered in Section 4.2.

We recognize that for this new Pareto–Pontryagin problem, i.e., to minimize (59) under the constraint (60), we have to specify some ingredients: the (perturbation of a) function φ will be offered by the biologists or agronomists denoting how the natural environment is linked with the standard description variables of the xylema. In other words, φ is resuming abundance or poverty of the chemical nutrients around the tree, together with lighting, temperature, humidity, etc. However, there is a very radically new element, the control parameter u , denoting the ability to choose and dose the quantities available. In other words, u represents the intelligent independent choices of the plant.

6. Conclusions and Future Research

In this paper, we have proposed a unification attempt of two research lines in applied optimization: Pareto multiobjective optimization and Pontryagin optimal control. We have reviewed the main features of the two theories, presenting the opportunity to reunify them by illustrating some typical problems in constrained optimization, some static application examples from Calculus of Variations, and dynamic applications of optimal control involving multiple objectives. All of these domains may benefit from a genuine Pareto

multiobjective reformulation of the theory of Pontryagin optimal control. With regards to this purpose, we have reformulated the Pontryagin maximum principle for a family of possibly conflicting functionals by merging the diverse functional in a linear convex combination, highlighting the role of the coefficients as Lagrange multipliers λ . By varying the choice of the multipliers λ , an infinite family of solutions emerges, which is the common case for the Pareto optimization. A possible natural extension of this theory consists of efficient numerical strategies for approximating the infinite family of solutions as a whole, as already proposed for the static case [25,29,41]. A natural continuation of the present work consists of a full dynamic implementation of the growth of xylems under physiological thrusts [33], as well as further applications in mechanical engineering [37] and materials science [30,31].

Author Contributions: Conceptualization, investigation and writing, A.L. and F.C. All authors have read and agreed to the published version of the manuscript.

Funding: A.L. acknowledges the funding from MIUR-Progetti di Ricerca di Interesse Nazionale “Mathematics of active materials: from mechanobiology to smart devices” grant number 2017KL4EF3. The APC was funded by University of Padova-Mathematics Department.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A. Lagrangian Multipliers and the Fundamental Homomorphism Theorem

Let us consider the set of maps

$$v(\cdot) \in \Gamma^V := \left\{ v(\cdot) \in C^\infty(\Omega^{d+1}; \mathbb{R}^M) : v \Big|_{\partial\Omega^{d+1}} = V \right\} \quad (\text{A1})$$

where $V : \partial\Omega^{d+1} \rightarrow \mathbb{R}^M$ are boundary conditions (constraints). Let us consider the variational principle

$$d\mathcal{J} = 0, \quad \text{where } \mathcal{J} : \Gamma^V \rightarrow \mathbb{R}, \quad v(\cdot) \mapsto \mathcal{J}[v] = \int_{\Omega^{d+1}} L(x, v(x), \nabla v(x)) dx. \quad (\text{A2})$$

The Gateaux derivative of $\mathcal{J}$ reads

$$d\mathcal{J}[v]\delta v = - \int_{\Omega^{d+1}} \left(\frac{\partial}{\partial x^i} \left(\frac{\partial L}{\partial v^{\Delta}_{,i}} \right) - \frac{\partial L}{\partial v^{\Delta}} \right) \delta v^{\Delta}(x) dx, \quad (\text{A3})$$

where

$$\delta v(\cdot) \in \Gamma^0 := \left\{ v(\cdot) \in C^\infty(\Omega^{d+1}; \mathbb{R}^M) : v \Big|_{\partial\Omega^{d+1}} = 0 \right\}. \quad (\text{A4})$$

Appendix A.1. Functional Constraints

Take into account k functional constraints $\mathcal{V}_\alpha|_{\alpha=1,\dots,k}$,

$$\mathcal{V}_\alpha : \quad \mathcal{J}_\alpha = c_\alpha, \quad \alpha = 1, \dots, k \quad (\text{A5})$$

$$\mathcal{J}_\alpha : \Gamma^V \rightarrow \mathbb{R}, \quad v(\cdot) \mapsto \int_{\Omega^{d+1}} L^\alpha(x, v(x), \nabla v(x)) dx. \quad (\text{A6})$$

Let $v^* \in \Gamma^V \cap \mathcal{V}$. Observe that

$$T_{v^*}\Gamma^V = \Gamma^0 := \left\{ v(\cdot) \in C^\infty(\Omega^{d+1}; \mathbb{R}^M) : v \Big|_{\partial\Omega^{d+1}} = 0 \right\} \quad (\text{A7})$$

and

$$T_{v^*}(\Gamma^V \cap \mathcal{V}) = \ker d\mathcal{J}_\alpha|_{\alpha=1,\dots,k}[v^*] \quad (\text{A8})$$

From the analysis of the following diagrams

$$\begin{array}{ccc}
 \Gamma^0 & \xrightarrow{d\mathcal{J}[v^*]} & \mathbb{R} \\
 \downarrow d\mathcal{J}_\alpha[v^*]|_{\alpha=1,\dots,k} & \nearrow \Lambda & \\
 \mathbb{R}^k & &
 \end{array}
 \quad
 \begin{array}{ccc}
 \delta v & \xrightarrow{\quad} & d\mathcal{J}[v^*]\delta v \\
 \downarrow & \nearrow & \\
 d\mathcal{J}_\alpha[v^*]|_{\alpha=1,\dots,k}\delta v & &
 \end{array}
 \quad (A9)$$

we see that v^* is an extremal of $\mathcal{J}$ under the constraints $\mathcal{V}_\alpha|_{\alpha=1,\dots,k}$ if

$$\ker d\mathcal{J}_\alpha|_{\alpha=1,\dots,k}[v^*] \subseteq \ker d\mathcal{J}[v^*]. \quad (A10)$$

The fundamental homomorphism theorem gives us

$$d\mathcal{J}[v^*] = \Lambda \cdot d\mathcal{J}_\alpha[v^*]|_{\alpha=1,\dots,k} \quad (A11)$$

In some more detail, writing

$$\Lambda = (\lambda_\alpha)|_{\alpha=1,\dots,k}, \quad (A12)$$

we have

$$\begin{aligned}
 0 &= (d\mathcal{J}[v^*] - \Lambda \cdot d\mathcal{J}_\alpha[v^*]|_{\alpha=1,\dots,k})\delta v \\
 &= \int_{\Omega^{d+1}} \left[\left(\frac{\partial L}{\partial v^{\Delta}_{,i}} \right) \delta v^{\Delta}_{,i} + \frac{\partial L}{\partial v^{\Delta}} \delta v^{\Delta} \right] - \lambda_\alpha \left[\left(\frac{\partial L^\alpha}{\partial v^{\Delta}_{,i}} \right) \delta v^{\Delta}_{,i} + \frac{\partial L^\alpha}{\partial v^{\Delta}} \delta v^{\Delta} \right] dx \\
 &= - \int_{\Omega^{d+1}} \left[\frac{\partial}{\partial x^i} \frac{\partial(L - \lambda_\alpha L^\alpha)}{\partial v^{\Delta}_{,i}} - \frac{\partial(L - \lambda_\alpha L^\alpha)}{\partial v^{\Delta}} \right] \delta v^{\Delta} dx. \quad (A13)
 \end{aligned}$$

Thus, if $v^* \in \Gamma^V \cap \mathcal{V}$ is a stationary point for $\mathcal{J}$ along variations which are tangents to the constraints, i.e., $\delta v \in T_{v^*}(\Gamma^V \cap \mathcal{V})$, then there exists $\Lambda \in \mathbb{R}^k$, depending on v^* , and a new functional $\overline{\mathcal{J}}$ which is stationary at v^* for any (unconstrained) variations $\delta v \in \Gamma^0$,

$$\overline{\mathcal{J}} := \int_{\Omega^{d+1}} (L - \lambda_\alpha L^\alpha) dx, \quad (A14)$$

where

$$\mathcal{L} = L - \lambda_\alpha L^\alpha \quad (A15)$$

is said to be the *augmented Lagrangian function*—compared with the combined Lagrangian functions in (29) and in (41) above.

Appendix A.2. Punctual Constraints

Whenever the new k constraints are *punctual* ones,

$$\varphi^\alpha(x) = c_\alpha, \quad \alpha = 1, \dots, k \quad \varphi^\alpha|_{\alpha=1,\dots,k} \in C^\infty(\Omega^{d+1}; \mathbb{R}^k) \quad (A16)$$

the above construction has to be slightly modified. We arrive at the following diagrams

$$\begin{array}{ccc}
 \Gamma^0 & \xrightarrow{d\mathcal{J}[v^*]} & \mathbb{R} \\
 \downarrow d\varphi[v^*] & \nearrow \Lambda & \\
 C^\infty(\Omega^{d+1}; \mathbb{R}^k) & &
 \end{array}
 \quad
 \begin{array}{ccc}
 \delta v & \xrightarrow{\quad} & d\mathcal{J}[v^*]\delta v \\
 \downarrow & \nearrow & \\
 d\varphi[v^*]\delta v & &
 \end{array}
 \quad (A17)$$

and finally the related new *augmented Lagrangian function* now reads

$$\mathcal{L} = L(x, v, \nabla v) - \lambda_\alpha(x) L^\alpha(x, v, \nabla v) \quad (\text{A18})$$

Unlike the previous case, now λ_α is a function and no longer a set of real numbers.

Appendix B. A Symplectic Framework for Optimal Control Theory

A careful reconnaissance of the deduction in Section 2 technically shows two important drawbacks.

First, we involved a singular Legendre transformation.

Second, the well-known decisive role in the optimality of u control variables has been relegated in the above reconstruction to a banal parametric dependence.

Recall that the PMP is telling us that if

$$\begin{aligned} (x(\cdot), u(\cdot)) \quad \text{minimizes} \quad & \int_0^T L(x(t), u(t)) dt, \\ \text{under the constraint} \quad & \dot{x} = \varphi(x, u), \quad x(0) = x_0, \end{aligned} \quad (\text{A19})$$

then

$$\mathcal{H}(x(t), p(t); u(t)) \geq \mathcal{H}(x(t), p(t); \omega), \quad \text{for all } t \in [0, T] \text{ and } \omega \in \text{Dom}(u). \quad (\text{A20})$$

This last aspect, in a full smooth environment, implies necessarily that

$$\frac{\partial \mathcal{H}}{\partial u}(x, p, u) = 0 \quad (\text{A21})$$

Here below we will see that the standard Hamiltonian equations

$$\dot{x} = \frac{\partial \mathcal{H}}{\partial p}(x, p, u), \quad \dot{p} = -\frac{\partial \mathcal{H}}{\partial x}(x, p, u), \quad (\text{A22})$$

together with the (A21), are intrinsically encoded into the following generalized Hamiltonian symplectic setting, see (A29).

Symplectic PMP

We consider a base manifold Q . To generalize the Hamiltonian vector fields $X_H : T^*Q \rightarrow TT^*Q$, we consider their image into TT^*Q and try to interpret them as suitable Lagrangian submanifolds.

The tangent bundle of the co-tangent bundle T^*Q , that is TT^*Q ,

$$(x, p, \dot{x}, \dot{p}) \in TT^*Q, \quad \dim TT^*Q = 4n, \quad (\text{A23})$$

becomes a symplectic manifold if, e.g., we endowed it with the following closed and nondegenerate 2-form Θ :

$$\Theta = d(\dot{p}dx - \dot{x}dp) = d\dot{p} \wedge dx - d\dot{x} \wedge dp. \quad (\text{A24})$$

Consider here the Lagrangian submanifolds. The frame is the following:

$$\begin{array}{ccccc} \Lambda & \xhookrightarrow{j} & TT^*Q & \xrightarrow{\tau} & T^*Q \\ \ell & \mapsto & (x(\ell), p(\ell), \dot{x}(\ell), \dot{p}(\ell)) & \mapsto & (x(\ell), p(\ell)), \end{array} \quad (\text{A25})$$

As before, a submanifold $\Lambda \subset TT^*Q$ is Lagrangian iff $\dim \Lambda = 2n$, and the restriction of Θ on Λ is vanishing. As announced, the image of a Hamiltonian vector field X_H

$$X_H : T^*Q \longrightarrow TT^*Q, \quad (x, p) \mapsto (x, p, \frac{\partial H}{\partial p}(x, p), -\frac{\partial H}{\partial x}(x, p)) \quad (\text{A26})$$

is Lagrangian. Effectively, the dimension is obviously one, and we see that

$$\begin{aligned} \Theta|_{\text{image}(X_H)} &= j^*\Theta = j^*d(\dot{p}dx - \dot{x}dp) = dj^*(\dot{p}dx - \dot{x}dp) = \\ &= d(-dH(x, p)) = -d^2H(x, p) = 0. \end{aligned} \quad (\text{A27})$$

Here the role of *generating function* is played by the Hamiltonian function.

A natural re-setting of the Maslov–Hörmander—see e.g., [42]—theorem at the actual level leads us to characterize locally the Lagrangian submanifolds of TT^*Q as the loci of the points $(x, p, \dot{x}, \dot{p})$ such that, for some function

$$H : T^*Q \times \mathbb{R}^k \longrightarrow \mathbb{R}, \quad (x^i, p_j, \xi^A) \mapsto H(x^i, p_j, \xi^A), \quad (\text{A28})$$

we have:

$$\dot{x}^i = \frac{\partial H}{\partial p_i}(x, p, \xi), \quad \dot{p}_j = -\frac{\partial H}{\partial x^j}(x, p, \xi), \quad 0 = \frac{\partial H}{\partial \xi^A}(x, p, \xi), \quad (\text{A29})$$

with a suitable rank condition on the second derivatives. The above Equation (A29) are exactly the Hamiltonian Optimal Control equations given by the *Pontryagin Maximum Principle*—see Sussmann’s papers, Formula (19) p. 39 in [43] and (V.12) p. 107 in [44]. In this symplectic framework, they are natural, giving us the more general Hamiltonian system structure.

This order of ideas arose from Tulczyjew. He has been able to give a coherent exposition of relativistic particle motion and to construct a very general version of the Legendre Transformation [45–47].

References

1. Edgeworth, F.Y. *Mathematical Psychics: An Essay on the Application of Mathematics to the Moral Sciences*; London: Kegan Paul; McMaster University Archive for the History of Economic Thought: Hamilton, ON, Canada, 1881.
2. Pareto, V. *Cours D’économie Politique/Professé à L’université de Lausanne*; Rouge: Lausanne, Switzerland, 1896.
3. Pareto, V. *Manuale di Economia Politica con una Introduzione alla Scienza Sociale*; Piccola Biblioteca Scientifica, Società Editrice Libreria: Milan, Italy, 1906.
4. Gambier, A.; Badreddin, E. Multi-objective Optimal Control: An Overview. In Proceedings of the 2007 IEEE International Conference on Control Applications, Singapore, 1–3 October 2007; pp. 170–175, ISSN 1085-1992. [CrossRef]
5. Peitz, S.; Ober-Blöbaum, S.; Dellnitz, M. Multiobjective optimal control methods for the Navier-Stokes equations using reduced order modeling. *Acta Appl. Math.* **2019**, *161*, 171–199. [CrossRef]
6. Peitz, S.; Dellnitz, M. A Survey of Recent Trends in Multiobjective Optimal Control—Surrogate Models, Feedback Control and Objective Reduction. *Math. Comput. Appl.* **2018**, *23*, 30. [CrossRef]
7. Visetti, D.; Heyde, F. Euler-Lagrange equations for multiobjective calculus of variations problems via set optimization. *arXiv* **2021**, arXiv: 1911.11754.
8. Zhu, Q.J. Hamiltonian Necessary Conditions for a Multiobjective Optimal Control Problem with Endpoint Constraints. *SIAM J. Control Optim.* **2000**, *39*, 97–112. [CrossRef]
9. Degiovanni, M.; Lucchetti, R.; Ribarska, N. Critical point theory for vector valued functions. *J. Convex Anal.* **2002**, *9*, 415–428.
10. Ngo, T.N.; Hayek, N. Necessary conditions of Pareto optimality for multiobjective optimal control problems under constraints. *Optimization* **2017**, *66*, 149–177. [CrossRef]
11. Alexéev, V.; Tikhomirov, V.; Fomine, S. *Commande Optimale*; MIR: Moscow, Russia, 1982.
12. Pontriaguine, L.; Boltianski, V.; Gamkrelidze, R.; Michtchenko, E. *Théorie Mathématique des Processus Optimaux*; MIR: Paris, France, 1974.
13. Smale, S. Global Analysis and Economics. I. Pareto Optimum and a Generalization of Morse Theory. In *Dynamical Systems (Proc. Sympos., Univ. Bahia, Salvador, 1971)*; Academic Press: New York, NY, USA, 1973; pp. 531–544.

14. Smale, S. Global analysis and economics III: Pareto Optima and price equilibria. *J. Math. Econ.* **1974**, *1*. [\[CrossRef\]](#)
15. Kuhn, H.W.; Tucker, A.W. Nonlinear programming. In Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability, Berkeley, CA, USA, 31 July–12 August 1950; University of California Press: Los Angeles, CA, USA, 1951; pp. 481–492.
16. Miettinen, K. *Nonlinear Multiobjective Optimization*; International Series in Operations Research; Springer: Boston, MA, USA, 1998; ISSN 0884-8289. [\[CrossRef\]](#)
17. Miglierina, E.; Molho, E.; Rocca, M. Critical points index for vector functions and vector optimization. *J. Optim. Theory Appl.* **2008**, *138*, 479–496. [\[CrossRef\]](#)
18. Miglierina, E.; Molho, E.; Rocca, M. *A Morse-Type Index for Critical Points of Vector Functions*; Technical Report 2007/02; Department of Economics, University of Insubria: Varese, Italy, 2007.
19. Miglierina, E. Characterization of solutions of multiobjective optimization problem. *Rend. Circ. Mat. Palermo Ser. II* **2001**, *50*, 153–164. [\[CrossRef\]](#)
20. Agrachev, A.A.; Sachkov, Y. *Control Theory from the Geometric Viewpoint*; Springer Science & Business Media: Berlin/Heidelberg, Germany, 2013.
21. Bressan, A.; Piccoli, B. *Introduction to the Mathematical Theory of Control*; AIMS Series on Applied Mathematics; American Institute of Mathematical Sciences (AIMS): Springfield, MO, USA, 2007; Volume 2.
22. Arnold, V.I.; Kozlov, V.V.; Neishtadt, A.I. *Mathematical Aspects of Classical and Celestial Mechanics*, 3rd ed.; Encyclopaedia of Mathematical Sciences; Springer: Berlin, Germany, 2006; Volume 3. [\[CrossRef\]](#)
23. Marsden, J.E.; Ratiu, T.S. *Introduction to Mechanics and Symmetry*, 2nd ed.; Texts in Applied Mathematics; Springer: New York, NY, USA, 1999; Volume 17. [\[CrossRef\]](#)
24. Miglierina, E.; Molho, E. Scalarization and stability in vector optimization. *J. Optim. Theory Appl.* **2002**, *114*, 657–670. [\[CrossRef\]](#)
25. Gebken, B.; Peitz, S. Inverse multiobjective optimization: Inferring decision criteria from data. *J. Glob. Optim.* **2021**, *80*, 3–29. [\[CrossRef\]](#)
26. Hartikainen, M.E.; Lovison, A. PAINT–SiCon: Constructing consistent parametric representations of Pareto sets in nonconvex multiobjective optimization. *J. Glob. Optim.* **2015**, *62*, 243–261. [\[CrossRef\]](#)
27. Lovison, A. Global search perspectives for multiobjective optimization. *J. Glob. Optim.* **2013**, *57*, 385–398. [\[CrossRef\]](#)
28. Lovison, A.; Pecci, F. Hierarchical stratification of Pareto sets. *arXiv* **2014**, arXiv: 1407.1755.
29. Lovison, A. Singular Continuation: Generating Piecewise Linear Approximations to Pareto Sets via Global Analysis. *SIAM J. Optim.* **2011**, *21*, 463–490. [\[CrossRef\]](#)
30. Truskinovsky, L.; Zanzotto, G. Ericksen’s bar revisited: Energy wiggles. *J. Mech. Phys. Solids* **1996**, *44*, 1371–1408. [\[CrossRef\]](#)
31. Truskinovsky, L.; Zanzotto, G. Finite-scale microstructures and metastability in one-dimensional elasticity. *Meccanica* **1995**, *30*, 577–589. [\[CrossRef\]](#)
32. Ericksen, J.L. Equilibrium of bars. *J. Elast.* **1975**, *5*, 191–201. [\[CrossRef\]](#)
33. Koçillari, L.; Olson, M.E.; Suweis, S.; Rocha, R.P.; Lovison, A.; Cardin, F.; Dawson, T.E.; Echeverría, A.; Fajardo, A.; Lechthaler, S.; et al. The Widened Pipe Model of plant hydraulic evolution. *Proc. Natl. Acad. Sci. USA* **2021**, *118*, e2100314118. [\[CrossRef\]](#)
34. Noor, E.; Milo, R. Efficiency in Evolutionary Trade-Offs. *Science* **2012**, *336*, 1114–1115. [\[CrossRef\]](#)
35. Shoal, O.; Sheftel, H.; Shinar, G.; Hart, Y.; Ramote, O.; Mayo, A.; Dekel, E.; Kavanagh, K.; Alon, U. Evolutionary Trade-Offs, Pareto Optimality, and the Geometry of Phenotype Space. *Science* **2012**, *336*, 1157–1160. [\[CrossRef\]](#) [\[PubMed\]](#)
36. Agrachev, A.A.; Sarychev, A.V. On abnormal extremals for Lagrange variational problems. *J. Math. Syst. Estim. Control* **1995**, *5*, 31.
37. Bertolazzi, E.; Biral, F.; Da Lio, M. Symbolic-numeric indirect method for solving optimal control problems for large multibody systems. *Multibody Syst. Dyn.* **2005**, *13*, 233–252. [\[CrossRef\]](#)
38. Biral, F.; Bertolazzi, E.; Bosetti, P. Notes on Numerical Methods for Solving Optimal Control Problems. *IEEE J. Ind. Appl.* **2016**, *5*, 154–166. [\[CrossRef\]](#)
39. Cossalter, V.; Da Lio, M.; Biral, F.; Fabbri, L. Evaluation of Motorcycle Maneuverability With the Optimal Maneuver Method. *SAE Trans.* **1998**, *107*, 2512–2518. [\[CrossRef\]](#)
40. Cossalter, V.; Da Lio, M.; Lot, R.; Fabbri, L. A general method for the evaluation of vehicle manoeuvrability with special emphasis on motorcycles. *Veh. Syst. Dyn.* **1999**, *31*, 113–135. [\[CrossRef\]](#)
41. Lovison, A.; Miettinen, K. On the Extension of the DIRECT Algorithm to Multiple Objectives. *J. Glob. Optim.* **2021**, *79*, 387–412. [\[CrossRef\]](#)
42. Cardin, F. *Elementary Symplectic Topology and Mechanics*; Lecture Notes of the Unione Matematica Italiana; Springer: Cham, Switzerland, 2015; Volume 16. [\[CrossRef\]](#)
43. Sussmann, H.J.; Willems, J.C. 300 years of optimal control: From the brachistochrone to the maximum principle. *IEEE Control Syst. Mag.* **1997**, *17*, 32–44. [\[CrossRef\]](#)
44. Sussmann, H.J.; Willems, J.C. *Three Centuries of Curve Minimization: From the Brachistochrone to Modern Optimal Control Theory*; 2003. Available online: <https://www.math.rutgers.edu/~sussmann/papers/main-draft.ps.gz> (accessed on 5 May 2022).
45. Menzio, M.R.; Tulczyjew, W.M. Infinitesimal symplectic relations and generalized Hamiltonian dynamics. *Ann. Inst. Henri Poincaré Sect. A* **1978**, *28*, 349–367.

-
46. Tulczyjew, W. A symplectic formulation of relativistic particle dynamics. *Acta Phys. Pol. Ser. B* **1977**, *8*, 431–447.
 47. Tulczyjew, W.M. Geometric formulations of physical theories. In *Monographs and Textbooks in Physical Science; Lecture Notes*; Bibliopolis: Naples, Italy, 1989; Volume 11.