Multi-objective optimisation of aircraft flight trajectories in the ATM and avionics context

Gardi, Alessandro; Sabatini, Roberto; Ramasamy, Subramanian

https://doi.org/10.1016/j.paerosci.2015.11.006

Document Version: Accepted Manuscript

Published Version: https://doi.org/10.1016/j.paerosci.2015.11.006

Repository homepage: https://researchrepository.rmit.edu.au

© 2016 Elsevier Ltd.
Downloaded On 2023/09/15 21:12:37 +1000
Thank you for downloading this document from the RMIT Research Repository.

The RMIT Research Repository is an open access database showcasing the research outputs of RMIT University researchers.


Citation:


See this record in the RMIT Research Repository at:


Version: Accepted Manuscript

Copyright Statement: © 2016 Elsevier Ltd
Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License.

Link to Published Version:

http://dx.doi.org/10.1016/j.paerosci.2015.11.006

PLEASE DO NOT REMOVE THIS PAGE
Multi-Objective Optimisation of Aircraft Flight Trajectories in the ATM and Avionics Context

Alessandro Gardi, Roberto Sabatini* and Subramanian Ramasamy

RMIT University, Melbourne, Victoria 3001, Australia

ABSTRACT

The continuous increase of air transport demand worldwide and the push for a more economically viable and environmentally sustainable aviation are driving significant evolutions of aircraft, airspace and airport systems design and operations. Although extensive research has been performed on the optimisation of aircraft trajectories and very efficient algorithms were widely adopted for the optimisation of vertical flight profiles, it is only in the last few years that higher levels of integration were proposed for automated flight planning and re-routing functionalities of innovative Communication Navigation and Surveillance/Air Traffic Management (CNS/ATM) and Avionics (CNS+A) systems. In this context, the implementation of additional environmental targets and of multiple operational constraints introduces the need to efficiently deal with multiple objectives as part of the trajectory optimisation algorithm. This article provides a comprehensive review of Multi-Objective Trajectory Optimisation (MOTO) techniques for transport aircraft flight operations, with a special focus on the recent advances introduced in the CNS+A research context. In the first section, a brief introduction is given, together with an overview of the main international research initiatives where this topic has been studied, and the problem statement is given. The second section introduces the mathematical formulation and the third section reviews the numerical solution techniques, including discretisation and optimisation methods for the specific problem formulated. The fourth section summarises the strategies to articulate the preferences and to select optimal trajectories when multiple conflicting objectives are introduced. The fifth section introduces a number of models defining the optimality criteria and constraints typically adopted in MOTO studies, including fuel consumption, air pollutant and noise emissions, operational costs, condensation trails, current airspace and airport operations. A brief overview of atmospheric and weather modelling is also included. Key equations describing the optimality criteria are presented, with a focus on the latest advancements in the respective application areas. In the sixth section, a number of MOTO implementations in the CNS+A systems context are mentioned with relevant simulation case studies addressing different operational tasks. The final section draws some conclusions and outlines guidelines for future research on MOTO and associated CNS+A system implementations.

KEYWORDS

4-Dimensional Flight Trajectory, Air Traffic Management, Aircraft Emissions, Avionics, Environmental Impacts, Flight Planning, Multi-Objective Optimisation, Optimal Control, Sustainable Aviation, Trajectory Optimisation

1. Introduction

The continuous growth of civil air transport and the increasing adoption of manned and unmanned aerial vehicles for new and more traditional roles is posing significant challenges to the aviation community, as the current paradigms will not ensure the desired levels of safety, efficiency and environmental sustainability in the future unless substantial evolutions are introduced. As a consequence, several large-scale aviation renovation initiatives were launched in the last two decades. These research and development activities are investigating, in particular, the most promising technological and operational improvements that would enhance the levels of safety, capacity, efficiency and environmental sustainability associated with current and likely future aviation business models in a holistic manner, hence by specifically improving the design, manufacturing, operation and lifecycle management of aircraft. In the operational domain, significant progresses in terms of safety, capacity and efficiency of air traffic are particularly expected from the implementation of novel concepts and technologies in the Air Traffic Management (ATM) and avionics domain, in line with the evolutions originally envisioned by the Future Air Navigation Systems (FANS) special committee of the International Civil Aviation Organization (ICAO) in the 1980s [1]. The ATM relies on a large set of operational measures to fulfil its mission of preventing collisions and promoting an ordered and expedite flow of air traffic [2, 3]. In the current centralised command and control-oriented ATM configuration, these measures are typically based on amending the lateral, vertical and longitudinal navigation of aircraft as necessary, and can be performed at various operational timeframes. As depicted in Figure 1, these measures can only be ensured if suitable means are available to identify the conflicts, negotiate a resolution and follow an alternative flight trajectory. For instrument flights in controlled airspace these essential functions are provided by Communication, Navigation, Surveillance, ATM (CNS/ATM)
and Avionics (CNS+A) technologies. Human operators are still responsible for the majority of duties involved in the current ATM paradigm, which is largely procedural and hence more tactical than strategic in nature [4]. The Civil Air Navigation Services Organisation (CANSO) estimated that global ATM is currently between 92-94% fuel-efficient and that unrecoverable factors limit the attainable efficiency improvements to between 95-98% [5]. The International Air Transport Association (IATA) quotes figures of 12% fuel inefficiency in relation to ATM [6]. Efficiency improvements of individual ATM procedures implemented locally are thereby considered largely ineffective, especially in view of the foreseen steady traffic growth trends. In order to achieve the ambitious medium-long-term goals set for safety, efficiency and environmental sustainability, substantial efficiency advances in the exploitation of the available airspace and airport resources are required.

Operational enhancements also have recognised potential benefits in mitigating environmental and health impacts of air transport, and effectively a number of research activities in the last decades have explored these potential benefits. Some operational measures are potentially more effective than aircraft design enhancements, such as in the case of perceived noise and contrails. While some study led to the successful implementation of procedural improvements such as enhanced instrument departure and arrival routings, the effectiveness of these measures in the bigger picture is limited, especially due to their offline nature that rigidly constrains the flexibility of operations and does not cope efficiently with unforeseen airspace blockages. Therefore, research on CNS+A system implementations of optimised arrival and departure routing algorithms, which is still under way, will be crucial in delivering the required flexibility and versatility to facilitate online reconfigurations.

1.1. ATM and avionics modernisation

CNS+A technologies and operational concepts are expected to deliver some of the most significant contributions to enhancing the operational performances of air transport system. The growing political and industrial support around the evolutions initially envisioned by the FANS special committee and subsequent proposals ultimately led to the launch of major programmes such as the Single European Sky ATM Research (SESAR) in Europe and of the Next Generation Air Transportation System (NextGen) in the United States, as well as a number of other related initiatives around the globe. NextGen is a major collaborative research initiative pursuing ATM modernisation, operational improvements, technology maturation, scientific knowledge and integrated modelling, alternative jet fuels and targeted policy measures [8]. SESAR and the Clean Sky Joint Technological Initiative (JTI) for Aeronautics and Air Transport (Clean Sky) are the leading programmes outlining the future of air transportation in Europe by addressing both operational improvements and environmental issues [9-11]. These initiatives were established with the particular aim of governing and financially sustain the aviation modernisation in a harmonised framework. The Collaborative Actions for Renovation of Air Traffic Systems (CARATS) is the main ATM modernisation programme in Japan, which has entered its implementation phase after the long-term vision and roadmap were defined in 2010 [12]. Comprehensive, these major international programmes support the evolution of ATM into a highly automated, integrated and more collaborative system, allowing a more flexible and efficient management of the airspace and airport resources through higher levels of automation and more accurate navigation to maximise the exploitation of the available capacity. The ICAO’s recently established Aviation System Block Upgrades (ASBU) framework builds upon these major air navigation improvement programmes, aiming at the harmonised evolution of the CNS+A systems capabilities. An ASBU block consists of several modules, each relating targeted operational improvements with the governing standards, procedures, technology and equipage required to implement them. The initiatives led by the National Aeronautics and Space Administration (NASA) in the United States resulted in creation of the Environmentally Responsible Aviation (ERA) project to explore and document the feasibility, benefits and technical risks of advanced vehicle concepts and technologies that will reduce the impacts of aviation on the environment. The outcomes of such initiatives have already resulted in improved fuel efficiencies, reduced noxious gas emissions and noise levels [9, 13]. In the Asia-Pacific region, a number of Air Navigation Service Providers (ANSP) formed the Asia and Pacific Initiative to Reduce Emissions (ASPIRE) in 2008, with the aim of trialling enhanced operational procedures. ASPIRE implements a pragmatic approach of targeting city-pairs where several “best practice” green procedures can be applied in concert to realise cumulative benefits. The ASPIRE Strategic Plan relates each ASPIRE “best practice” green procedure to an ASBU module. In Australia, optimised ATM procedures such as tailored arrivals [14] and the green Required Navigation Performance (RNP) project [15] have been implemented. These national initiatives are aligned with those of the Asia-Pacific region with Australia’s involvement in the ASPIRE programme. Optimised Flextracks have been also established, allowing long-haul traffic to maximise the benefit offered by favourable winds in the relative low traffic densities encountered in the southern hemisphere. Soliciting aircraft equipage evolutions is one of the key ASBU objectives, but the relative youth of fleets in the fastest growing markets such as Asia-Pacific and South-America ensures that advanced CNS+A equipage is more commonly available where it is mostly needed in the future. A significant number of concepts were proposed for implementation as part of the major ATM modernisation programmes. Figure 2 depicts the concepts selected for implementation as part of the SESAR ATM master plan [16]. The measures are classified based on their relevance for en-route operations, Terminal Manoeuvring Area (TMA) operations, airport operations and on the network. Green boxes represent Pilot Common Project (PCP) essential operational changes and blue boxes represent new essential operational changes. Increasingly higher amounts of information will be collected, analysed and shared among the ground-based ATM and the airborne avionic systems to more effectively deal with unpredicted events and mitigate disruptions. In order to optimally exploit these quickly growing amounts of information made
available, an increase in automation support and a move away from the centralised command and control-oriented ATM paradigm towards more distributed/collaborative planning are deemed necessary. This will involve a redistribution of current ATM functions and services to other key players such as Airline Operation Centres (AOC) and flight crews, to improve the efficiency of the system as a whole.

The novel automated avionics and ATM DSS shall facilitate aircraft in flying user-preferred optimal flight paths (intsents), limiting the intervention of the air traffic controllers to high-level and emergency decisions, so to decrease the overall workload without compromising the situational awareness. In order to do so, these DSS are expected to dependably produce feasible trajectory solutions that resolve the impending conflicts, while complying with the timeframe allocated for the complete ATM transaction. The resulting increase in airspace flexibility will enable greener and more efficient flight profiles.

In order to be economically and operationally viable, the novel CNS+A technologies and operational concepts must be developed and deployed in a phased manner. As depicted in figure 3, the stages for such evolution within SESAR were defined based on the capability and consist of [16]:

- **Time-based Operations**: for which strategic and tactical ATM and Air Traffic Flow Management (ATFM) actions are aimed at an optimal traffic synchronisation. The time of arrival of traffic at specific points is the fundamental metric being estimated, managed and monitored by all the involved entities both on ground and airborne.
- **Trajectory Based Operations (TBO)**: focussing on a further-evolved predictability, flexibility and environmental sustainability of air traffic, unleashing additional capacity. This stage involves the evolution of the legacy flight plans into dynamically managed 4DT, which become the continuously updated and negotiated reference plan for the aircraft mission.
- **Performance-Based Operations (PBO)**: for which all the available CNS performance is exploited to establish a high-performance, network-centric, collaborative, integrated and seamless ATM system, supporting high-density operations. In this stage, ATM services are customised depending on the highest level of CNS performance provided by the involved traffic, enabling a further enhanced exploitation of airspace capacity.

TBO are based on the adoption of 4DT defining the aircraft’s flight path in three spatial dimensions (i.e., latitude, longitude and altitude) and in time from origin to destination, and on the associated precise estimation and correction of current and predicted traffic positions [20]. Each aircraft is assigned a 4DT contract, which is determined by means of a CDM process involving novel ground-based and airborne DSS, evolving from the original reference business trajectory. Increased efficiency and higher throughput are obtained in a CNS+A context by actively managing 4DT. In the PBO context, the ATM services will be matched to the performance capability of aircraft. Airline deploying PBO-capable equipment will benefit easier access to congested areas and time periods. The regulations will impose requirements in terms of system performance rather than in terms of specific technology or equipment. Some of these CNS+A technologies are already approaching the market, while early stage advancements in the juridical framework are accommodating enhanced operational capabilities. Since most of the innovations currently being implemented were conceived from the operational point of view, the concurrent development of an adequate theoretical framework and the execution of extensive modelling and simulation activities are crucial.

In summary, the key CNS+A advances identified by the major aviation modernisation programmes around the globe include [17-19]:

- **Four Dimensional (4D) trajectory-based operations**;
- **Higher levels of Collaborative Decision Making (CDM)** to allow all involved parties to participate in the enhancement of system performance by sharing and accessing more accurate and updated information;
- **Role shifting of ATM from command and control oriented units to a highly automated and collaborative decision-maker in an interoperable network-centric environment**;
- **Dynamic Airspace Management (DAM) for an optimised exploitation of airspace capacity**;
- **Improved avionics and ATM systems Human Machine Interface and Interaction (HMI) design, interoperability and higher levels of automation**;
- **Performance-Based Communication, Navigation, Surveillance (PBC/PBN/PBS), enabling Performance-Based Operations (PBO)**.

In order to introduce these innovative concepts and ultimately progress along the planned evolutionary pathways, a number of new CNS+A technologies were considered essential, including:

- **Avionics and ATM Decision Support Systems (DSS)** featuring automation-assisted 4DT trajectory (4DT) planning and negotiation/validation functionalities;
- **Enhanced ground-based and satellite-based aeronautical communications systems**, involving a substantial exploitation of data-links;
- **Enhanced navigation accuracy and integrity by means of Ground-, Avionics- and Satellite-Based Augmentation Systems (GBAS/ABAS/SBAS)**, promoting Global Navigation Satellite Systems (GNSS) as primary means of navigation;
- **Enhanced ground-based and satellite-based surveillance**, including Automated Dependent Surveillance Broadcast (ADS-B) and self-separation;
- **A System Wide Information Management (SWIM) network**.

![Fig. 2. Concepts proposed for implementation within the SESAR masterplan [16].](image)
1.2. Environmental sustainability of aviation

The aviation industry is directly responsible of a number of adverse effects on the environment and on the living beings. Although ground support vehicles and infrastructure participate to the overall total impacts, the mitigation of impacts directly associated with the aircraft flight sees a considerable involvement by researchers and policy-makers worldwide. Figure 4 depicts the main environmental impacts associated with the various flight phases of typical transport mission. While some pollutants such as carbon dioxide ($CO_2$), unburnt hydrocarbons (HC), soot and sulphur oxides (SO$_x$) are understood to adversely affect the environment along the entire flight, others either do not always eventuate, as in the case of condensation trails (contrails), or their adverse effects are reputed significant mostly in proximity of the ground, as in the case of aircraft noise and carbon monoxide (CO). Nitrogen oxides (NO$_x$) in addition to their toxicity to the living beings at higher concentrations when emitted in proximity of the ground (tropospheric NO$_x$) are also known to trigger a family of chemical processes ultimately causing the depletion of the ozone layer, and are thereby associated with a positive radiative forcing.

The Advisory Council for Aviation Research and Innovation in Europe (ACARE) has set specific target goals to address environmental sustainability of aviation in its Strategic Research Agenda (SRA) and Strategic Research and Innovation Agenda (SRIA) [21, 22]. In Europe, air transportation causes approximately 3% of the total greenhouse gas emissions [23]. The Clean Sky programme implements the ACARE SRIA FlightPath 2050 objectives in multiple phases. Similarly to Clean Sky, other aviation modernisation programmes worldwide have set ambitious environmental objectives to reverse the adverse effects of air transport on the environment. Most programmes are attempting the achievement of the environmental targets by taking a holistic approach, hence encompassing the design, manufacturing, operation and lifecycle management of aircraft. In the operational domain, although the paradigm “less fuel consumed equates to less emissions” still meets the widest popular agreement and is aligned with the financial interests of aircraft operators, it is evident that a number of operational measures could be very effective in further mitigating environmental and health impacts, potentially more than design enhancements, such as in the case of perceived noise and contrails. On the global scale, the environment branch of the Air Transport Bureau (ATB) is in charge of the initiatives of the ICAO’s Committee on Aviation Environmental Protection (CAEP). The ICAO environmental branch periodically reviews the progresses achieved as well as the solutions proposed and adopted at national and international levels, leading to the publication of a comprehensive environmental report [24]. Current emphasis is on improving the accuracy of pollution data by enhancing the measurement and estimation methods. Figures 5 and 6 represent the current state-of-the-art characterisations of Local Air Quality (LAQ). The values are yearly averages.
main source of pollution figures, but the various neighbouring pollutants and the cumulative aerodynamic advection and diffusion of pollutants affect the accuracy of the measure and thus prevent discerning individual aircraft and specific phases of their flight. Pollutant emissions can also be empirically estimated based on commonly equipped thermodynamic and mechanical sensors that measure Fuel Flow (FF), Turbine Entry Temperature (TET), Exhaust Gas Temperature (EGT), Engine Pressure Ratio (EPR), engine rotation speeds \(N_1, N_2, \ldots\) and acoustic vibrations. These estimations rely upon empirical and semi-empirical models based on nominal engine and fuel conditions. Although typically adequate for Multi-Objective Trajectory Optimisation (MOTO) studies, these models show limited accuracy in estimating the emissions, as anomalies in the engine or fuel conditions can void key assumptions and compromise the validity of these model-based estimations. For more detail on the assessment of environmental impacts of aviation, the reader is referred to [28-34].

1.3. Early trajectory optimisation research in aviation

Most of the early research on trajectory optimisation in aviation aimed at improving the way aircraft were flown with respect to predefined air routes and instrument procedures, as these were the operational paradigm of the period. Improvements were sought in terms of lower fuel consumption and flight time, as these factors were directly correlated to Direct Operating Costs (DOC) sustained by aircraft operators/airline companies. Since the adopted Instrument Flight Rules (IFR) procedures and air routes rigidly constrained the lateral navigation, the early research mostly targeted optimised vertical planning (altitude and speed). Notable early endeavours with respect to optimised lateral planning targeted wind-optimal routing for offline transcontinental and trans-oceanic planning applications, mainly because of the higher freedom in routing across these areas. Due to their higher technicality these works are discussed in section 5.5.

A report published in the 1963 by the RAND Corporation discussed in detail the adoption of calculus of variation theory for solving aerospace trajectory optimisation problems, also including an approach based on Green’s Theorem [35]. The report was ultimately focussed on spacecraft trajectory optimisation problems, which encompassed a considerable portion of the research in the domain at that time as well as in the following decades. The report noted that since the theoretical notions and skills involved in the application of the variational calculus were not conventionally part of the body of knowledge of aerospace engineers, significant research efforts would have been required for the development of dedicated solutions. Some earlier works [36, 37] proposed some solution techniques customised for the optimisation of aerospace vehicle trajectories. Erzberger performed considerable research in the application of trajectory optimisation for transport aircraft [38-43]. In [38, 39] the energy state of the aircraft is adopted to analyse the minimum-fuel and minimum-time vertical trajectory profiles (i.e. altitude and airspeed) of short-haul aircraft. The evaluation is extended to the case involving horizontal wind. The study notes that neglecting the wind in the computation of optimal vertical profiles can be an acceptable approximation for short-haul flights due to the small deviation between these conditions, but the deterioration of performance would be considerable for long-haul flights. A subsequent work further detailed the findings in view of an operational adoption, thereby evaluating the approximations of constrained thrust and independency of fuel consumption upon the thrust [40]. The analysis identified that by including the dependency of fuel consumption on thrust, the trajectories optimised for non-constrained thrust differ noticeably from the constrained thrust ones. This consideration highlighted the opportunity of developing performance management systems providing the required computational capabilities on-board aircraft. The topic was very actively researched in the following period [44-49]. A review article published in 1979 concluded that fuel savings of 5% to 25% could be realised in different categories of aircraft by applying the reviewed trajectory optimisation techniques for vertical profile planning [41]. The study also noted that optimisation algorithms evaluating the opportunity to change cruise flight level based on wind profiles variations and subject to ATC constraints should be investigated. Further work noted the consisting differences between minimum-fuel and minimum-time trajectories in terms of climb and descent angles, cruise altitudes and airspeed [42]. Miele adopted vertical trajectory optimisation methods to study the guidance problem of flight recovery in the case of an aborted landing associated to windshear, both as horizontal shear and including a vertical downdraft component [50, 51]. The work highlighted the opportunity of developing advanced windshear control systems entailing take-off, aborted and penetration landings. The growing economic, and operational importance of optimised vertical flight planning and the good accuracy and computational efficiency of the algorithms made the aforementioned implementations highly successful in the following decades, and expectedly a large number of current generation Flight Management Systems (FMS) include such functionalities.

1.4. Trajectory optimisation research in the context of avionics and CNS/ATM evolutions

With the deployment of technological CNS advancements such as satellite navigation and data-links, and the concurrent introduction of enhanced operational paradigms, the last decades have been characterised by a renewed interest for trajectory optimisation in aviation. Visser [4] described in detail the application of trajectory optimisation in the context of forthcoming 4D time-based operations. He noted that advanced 4D guidance techniques are one of the cornerstones of the future ATM, especially in the Terminal Manoeuvring Area (TMA). The study considered two possible future ATM scenarios, which have become the mainstream paradigms for the subsequent research. In the more “democratic” one, the 4D-capable FMS would generate its preferred trajectory and negotiate it with the ground-based ATM system, which would subsequently perform a collective optimisation to resolve conflicts between individual trajectories and eventually uplink new constraints if the user-preferred trajectory incurs conflicts. The alternative “autocratic” approach involves a centralised decision-making process performed by the ground-based ATM system, which processes minimal information from the aircraft such as performance criterion and desired landing time, and broadcasts the conflict-free route clearances to all aircraft. Aircraft may only reject the 4D clearances if leading to performance violations or route inconsistencies. In both the democratic and autocratic scenarios adequate Air/ground Data Links (ADL) are essential, though it was estimated that the democratic approach would pose even higher requirements in terms of bandwidth and negotiation time than the autocratic one. The centralised approach was eventually selected in the conceptual investigation due to a number of reasons. The report reviewed the trajectory optimisation approaches proposed and implemented previously, noting that the success of the optimised vertical planning functionalities in FMS was largely due to the adoption of very efficient algorithms based on reduced-order energy-state dynamics in the vertical plane only, and aimed at reducing Direct Operating Costs (DOC) over a fixed
The study concludes that an approach based on direct collocation and nonlinear programming, involving suitable evolutions in the flight dynamics models and control techniques compared to the previous generation, is the most promising for developing a flexible real-time multi-aircraft optimised 4D trajectory optimisation algorithm. Such approach would still prove valid in case of ill-defined problems (i.e. involving unattainable constraints), by adopting a multi-objective formulation based on sequential goal programming. The study highlighted, nonetheless, the very considerable dependencies of the optimisation performances on the initial guess and on the recipe defining the distribution of collocation nodes.

Betts performed a considerable research on the implementation of direct transcription and collocation for atmosphere trajectory optimisation problems [52-61], for which the implementation of efficient sparse nonlinear programming was an essential aspect. Hagelauer and Mora-Camino investigated the application of dynamic programming and identified knowledge to limit the search space and state transitions, and by introducing neural networks it was possible to restrict the computation times considerably and enable future implementation for next generation FMS [62]. Yokoyama and Suzuki studied the application of genetic algorithms to identify a good initial guess before introducing a direct collocation solved by means of a Block Diagonal Hessian (BDH) method [63]. Jardin identified that some key issues in the development of real-time optimal deconflicted trajectory generation algorithms are associated with wind-optimal routing, conflict detection and optimal conflict resolution, and consequently proposed the adoption of a discrete dynamic programming approach featuring neighbouring optimal wind routing and deterministic conflict grid [64, 65]. The application of genetic algorithms for MOTO was discussed in [66]. Jacobsen and Ringertz studied the modelling of airspace constraints for horizontal trajectory optimisation, specifically addressing polygon-shaped constraints [67]. Elele and Richards proposed the adoption of a branch-and-bound step to efficiently solve optimal path planning problems involving obstacle avoidance or airspace constraints [68]. This allows splitting the horizontal path planning problem into a number of simpler subproblems, and potentially permits the identification of the global optimality region. Nine different branching strategies were considered. A multiphase hybrid trajectory optimisation case was studied in [69]. A Mixed-Integer Nonlinear Programming (MINLP) formulation was subsequently investigated in [70]. A customised multiphase optimal 4DT planner based on Hermite-Simpson collocation and interior-point nonlinear programming solvers was adopted to study the optimisation of multi-constrained 4DT in a polynomial interpolated 3D wind fields [71]. Clarke applied trajectory optimisation techniques for the development of continuous climb and descent profiles addressing multiple aircraft types in representative traffic flow scenarios, which are being implemented as standard departure and arrival procedures in a number of US airports [72-77].

A recent review of trajectory design models was performed by Delahaye et al. [78]. The source compares various alternatives available to discretise the flight trajectory of an aircraft, including straight segments, Lagrange polynomials, Hermite interpolation, piecewise linear, quadratic and cubic interpolations, conventional splines and basis splines, Bézier curves, principal component analysis and homotopy trajectories. Some algorithms that prove particularly suited for trajectory planning are also reviewed, including fast marching, ordered upwind and light propagation algorithms. Although trajectory optimisation with respect to multiple objectives is not the core issue tackled in the review, the lessons learned lead to identify the best optimisation strategies for strategic planning (i.e. offline strategic), pre-tactical and tactical planning (i.e. offline tactical and online), and emergency. Genetic algorithms are recommended for the strategic planning, in conjunction with the very large dimensions and complexity of the problem. An approach based on penalty volumes and light propagation algorithms is recommended for the pre-tactical and tactical stages, whereas optimal control is preferred for the emergency trajectory design.

1.5 Problem statement

On-board algorithms for generating optimum trajectories were considered computationally realisable as early as the 1970s [41], and effectively the following two decades have seen the implementation of a number of new automatic flight and flight director modes for most phases of flight, targeting in particular the vertical guidance and optimised standard turns. Although the savings derived from these measures were very significant with respect to the previous operational paradigms, it has been proposed that more substantial gains can be attained by adopting real-time optimisation strategies for strategic and tactical replanning of 4D flight trajectories in an intent-based network-centric ATM scenario. As previously mentioned, 4DT represent a substantial evolution when compared to conventional flight plans and their associated limitations. From a certain perspective, since flight plans are submitted offline as a substantially static entity, unforeseen weather and air traffic scenarios have the effect of progressively compromising their validity and optimality. Optimal 4DT planning and negotiation/validation are key features in novel avionics and ATM DSS designs, and introduce new variables and equations [4]. In addition to consistently and
reliably fulfilling the transaction time requirements for strategic and tactical ATM operations, the 4DT optimisation functionalities will have to integrate all the constraints and restrictions associated with air navigation rules, air traffic services and real traffic conditions in order to be ultimately implemented in airborne avionics and ground-based ATM DSS. The optimisation performed by current online DSS, such as FMS, is effectively still targeting the vertical flight profile only, as direct great circle arcs and constant-bank turns are implemented for lateral path planning. The generated route may be significantly suboptimal on a larger scale when considering wind field and airspace blockage due to traffic congestions, convective weather, volcanic ash and other localised phenomena. Furthermore, current DSS address the needs arising from operating in dense traffic and complex airspace by allowing the addition of various constraints, terms of airspeed, altitude and flight path. Constraints and route amendments are mostly introduced in relation to the manual intervention of human operators (i.e. flight crews and air traffic controllers), therefore their update frequency is limited, and this is known to affect the routing optimality when traffic or weather conditions are rapidly evolving. Additionally, the unconstrained optimal flight profile is not regenerated nor considered between various route amendments, and this lead to progressive deviations from optimality. It is known that the occurrence of unfeasible constraints requires the implementation of MOTO techniques such as sequential goal programming. Another limitation of current systems requiring the implementation of MOTO techniques is due to the very narrow set of optimality criteria and their limited modelling. In particular, current DSS only consider direct operating costs related to fuel consumption and flight time and do not take into account the wind field and airspace blockage due to traffic and atmospheric phenomena. The limited optimality set does not allow the mitigation of actual environmental impacts, since their dependencies on geographic location (e.g. perceived noise, contrails) and throttle settings (e.g. soot, unburned hydrocarbons, carbon monoxide, nitrogen oxides) are not captured by current algorithms. These limitations are increasingly evident with the growth of traffic. Increasing R&D initiatives are therefore addressing more effective and versatile mission and trajectory planning, optimisation and management functionalities incorporating MOTO algorithms to be implemented in novel airborne and ground-based CNS+A DSS. For instance, CNS+A implementations of MOTO are being explored as part of international research programmes such as the Systems for Green Operations Integrated Technology Demonstrator (SGO-ITD) of Clean Sky. To implement the envisioned improvements in terms of safety, capacity, efficiency and environmental sustainability of air traffic operations, it is essential to correlate the trajectory design parameters to the targeted benefits. This can be done by introducing suitable models, objectives and constraints in MOTO algorithms to plan the optimal 4D intents and at the same time estimate the environmental impacts and operational costs [19]. Figure 7 represents the concept of MOTO to tackle multiple operational, economic and environmental criteria in the aviation context. In order to obtain optimal 4D intents with respect to the set optimality objectives, the MOTO suite comprises a number of essential aircraft, environmental and operational models. These models include local/global weather, operational costs, pollutant emissions, airspace structure, contrails and aircraft noise. As the emphasis is on the noise perceived by the population on the ground the aircraft noise model must be complemented by suitable demographic distribution and digital terrain elevation databases.

Significant research activities and outcomes on multi-objective 4DT optimisation (MOTO-4D) algorithms are being achieved as part of the major research programmes worldwide. The progresses in optimal control theory and nonlinear programming

![Diagram](image-url)
are leading to considerable improvements to the numerical solution methods for real-time trajectory optimisation applications. A considerable opportunity for direct exploitation of these research outcomes into the next generation avionics and ATM systems is emerging. The aim of this article is to review the current scientific knowledge regarding the optimisation of transport aircraft flight trajectories with respect to multiple and typically conflicting objectives arising from the inclusion of multiple environmental and operational criteria, and to deduce or infer all the useful notions for the development of algorithms that are specifically conceived for the implementation in novel CNS+A systems. More specifically, section 2 introduces the general formulation of trajectory optimisation problems, with particular emphasis on the optimal control formulation. Sections 3, 4 and 5 summarise the theoretical framework and provide a conceptual map, enabling the reader to access further literature for specific implementations. More in detail, section 3 reviews the numerical solution strategies, section 4 introduces the theoretical framework and the decision logics associated with multi-objective optimality, and section 5 presents a number of supplementary models that are typically necessary for the definition of the optimality criteria and constraints in aeronautical applications of trajectory optimisation. Section 6 outlines the ongoing research on CNS+A DSS. Finally, Section 7 presents some conclusions.

2. Theoretical framework

In general, an optimisation process involves the adoption of suitable algorithms, decision logics and heuristics to identify the best element from a finite or infinite set of available alternatives. Trajectory optimisation studies the methods to determine the best possible trajectory of a dynamical system in a finite-dimensional manifold, in terms of specific objectives and adhering to given constraints and boundary conditions. By this very definition, a substantial similarity between the statements of the Trajectory Optimisation Problem (TOP) and of the Optimal Control Problem (OCP) can be noted. This comes as no surprise, as the assumed origin of both is the Brachistochrone problem, posed by Johann Bernoulli in 1696 [79]. Consequently, the two designations are frequently interchanged in the literature. One distinction noted is about the mathematical nature of the unknown solutions, as in OCP the interest is about the identification of the optimal input functions, whereas in TOP the search is typically restricted to a finite set of static input parameters [80]. This consideration partly justifies the success of TOP solution methods based on parameterisations. Furthermore, the formal distinction of control, state and output variables is not always implemented in TOP, possibly due to their early historical appearance and to the investigation of solution strategies beyond the framework of the optimal control theory, which will be briefly mentioned here. Notwithstanding, the continued familiarity with the OCP formulation was very fruitful, as the theoretical and computational advances in optimal control were directly exploited in TOP solution methods and more general in the aerospace vehicle Guidance, Navigation and Control (GNC) domain. After the early works on the solution of the Brachistochrone problem, further studies by Euler and Lagrange led to the establishment of the calculus of variations. Since, as mentioned, the sought optimal solutions are in the form of functions, the calculus of variations effectively refers to functional optimisation. As a consequence, an early developed approach to the solution of OCP that is still commonly considered at present is based on the calculus of variations. Conversely, the transcription of OCP in finite-dimensional Non-Linear Programming (NLP) problem is increasingly popular as a computationally fast OCP solution strategy. As we will expand later, most OCP solution methods can be ascribed to either of these approaches, and therefore are conventionally categorised as either direct methods, if based on the transcription to a finite NLP problem, and indirect methods, if theoretical derivations based on the calculus of variation are implemented to formulate a Boundary-Value-Problem (BVP), and both approaches have led to very successful numerical implementations [4, 61, 80, 81]. Such distinction is less clear-cut than originally claimed, as some methods either fall in between or are intentionally hybridised. Additionally, more recent studies are addressing the theoretical analogies between the two philosophies, and the findings are progressively overruling the distinction [80]. A third class of OCP solution strategies is represented by heuristic methods, such as simulated annealing, evolutionary algorithms, tree/graph/pattern search and particle-swarm.

Theoretical solution strategies for OCP, such as the ones developed exploiting NLP and the calculus of variation were highly instrumental for the development of numerical solution techniques for TOP. Notwithstanding, in principle TOP can also encompass alternative non OCP-based formulations, including linear/nonlinear parametric optimisation problems on both continuous and discrete search spaces. This is particularly noteworthy in the aviation context, due to the widespread reliance on a geometrical organisation for the airspace and air routes, upon which all flight trajectory descriptors have been based, as well as the rules of the air, piloting techniques, traffic separation criteria, and ATM lexicon. In particular, the current air navigation procedures, as well as the air route network, rely on discrete geographical descriptors for the horizontal flight path – known as Lateral NAVigation (LNAV) – and on altitude and airspeed constraints for the vertical flight profile – known as Vertical NAVigation (VNAV). Consequently, some studies in the operational aviation domain have approached the optimisation of flight trajectories within the current air navigation framework. This category of studies has been informally called procedural optimisation, as most frequently it involves the optimization of Standard Instrument Departure (SID) procedures and Standard Terminal Arrival Routes (STAR).

2.1. Optimal control problem

In view of the familiarity between the TOP and the OCP discussed previously, the most general and convenient way to formulate the TOP is based on the optimal control. Consequently, in line with the control theory, we introduce the vector of time-dependent state variables $x(t) \in \mathbb{R}^n$, the vector of time-dependent control variables $u(t) \in \mathbb{R}^m$, the vector of system parameters $p \in \mathbb{R}^q$ and the time $t \in [t_0; t_f]$. In the following subsections we introduce the dynamic constraints, the path constraints, the boundary conditions and the cost functions. A consistent definition of all these components is fundamental to formulate a well-posed OCP and to perform an appropriate selection of the numerical solution method and of the multi-objective technique.

2.2. Dynamic constraints

The specificity of the trajectory optimisation and optimal control with respect to other mathematical optimisation branches is the application to dynamical systems, i.e. in motion or transition along time. Therefore, a key component in the TOP formulation is the set of dynamic constraints, which are meant to reproduce the feasible motion of the system (i.e. the aircraft, in our case) within the TOP. A system of Differential Algebraic Equations (DAE), consisting of the time derivatives of the state variables, is usually
adopted to introduce the system dynamics, and the dynamic constraints are therefore written as:

\[ x(t) = f[x(t), u(t), t, p] \quad (1) \]

Nonlinear dynamics are natively encompassed, while other cases such as discrete-time dynamics may also be accounted for by adopting adequately relaxed formulations. The optimal control of discrete-time systems is beyond the scope of this article, and for more detail the reader may refer to [82-84].

2.3. Path constraints

In the generalised TOP formulation, all non-differential constraints insisting on the system between the initial and final conditions are classified as *path constraints*, as they restrict the path, i.e. the space of states and controls, of the dynamical system. In order to represent all possible non-differential restrictions on the vehicle motion, two types of algebraic path constraints are considered: *inequality constraints* and *equality constraints*. A generalised expression of an inequality constraint is:

\[ g_i(x(t), u(t), t; p) \leq 0 \quad (2) \]

whereas an equality constraint can be written as:

\[ h_i(x(t), u(t), t; p) = 0 \quad (3) \]

Equality constraints (eq. 3) can be considered a subset of inequality constraints, as they can be split in two opposite inequalities such as in:

\[
\begin{align*}
    g_{1a}(x(t), u(t), t; p) & \leq 0 \\
    g_{1b}(x(t), u(t), t; p) & \leq 0 \\
\end{align*}
\]

where \( g_{1a} = -g_{1b} \), hence eq. (2) is frequently the only one quoted. A compact representation of growing popularity due to a high similarity to the numerical implementation is [85]:

\[ c_{\text{min}} \leq c[x(t), u(t), t; p] \leq c_{\text{max}} \quad (5) \]

for which equality constraints are simply encompassed by imposing \( (c_{\text{min}})_i = (c_{\text{max}})_i \). Each of the \( i \)-th compound inequalities of eq. (5) can be split in two opposite inequality constraints similarly to eq. (4). Despite the generality of the unified formulations (eq. 2 and 5), it is typically preferable to treat equality constraints separately, as they introduce an additional relationship between some state and/or control variables, hence they usually allow the reduction of the number of states or controls in the TOP. Inequality constraints have a more ambiguous impact on the TOP, as sometimes the path of the dynamical system lies far from the constraint, and hence the latter could be safely ignored as it does not have any influence in the solution. Therefore, at each instant in the TOP time domain, an inequality constraint can either be *active*, if the system path intersects the constraint such that \( g_i(t) = 0 \) has to be enforced, or *inactive*, when \( g_i(t) < 0 \) is naturally verified, therefore the constraint can be safely ignored. As a consequence, for each inequality constraint the path of the dynamical system can be effectively subdivided in *constrained subarcs* and *unconstrained subarcs*. Unfortunately, this is computationally nontrivial: usually, the occurrence and the exact number of constrained subarcs along the path is not known a priori, as also very frequently is unknown the location of the junction points between unconstrained and constrained subarcs, or between two diversely constrained subarcs.

2.4. Boundary conditions

Boundary conditions define the values that state and control variables of the dynamical system shall have at the initial and final times. Since in some instances the boundary conditions are not always restricted to definite values, it is useful to adopt a generalised expression including relaxed conditions. Similarly to the already introduced expression of eq. (5) for the path constraints, we then write the boundary conditions as:

\[ B_{\text{min}} \leq B[x(t_0), x(t_f), u(t_0), u(t_f); p] \leq B_{\text{max}} \quad (6) \]

where equality conditions are still encompassed by imposing \( (B_{\text{min}})_i = (B_{\text{max}})_i \).

2.5. Cost functions and performance indexes

In order to optimise a given performance, it is necessary to introduce a scalar index, the *performance index*, which by means of a suitably defined *cost function* quantifies the achievement of that particular objective. The optimisation process can then be translated into the mathematical minimisation (or maximisation) of such performance index. Generally speaking, a performance index may depend on a function of terminal state values and parameters, \( \Phi[x(t_f), u(t_f), p] \), or on the integral of one or multiple state variables, control variables, and parameters, that is \( \int_{t_0}^{t_f} \Psi[x(t), u(t), p] dt \). The generic TOP formulation involving a performance index \( J_i \) that takes into account both components was introduced by Bolza [79, 86, 87], and is expressed as:

\[ J_i = \Phi[x(t_f), u(t_f), p] + \int_{t_0}^{t_f} \Psi[x(t), u(t), p] dt \quad (7) \]

The optimisation is classified as *single-objective* when an individual performance index \( J \) is introduced and *multi-objective* when two or multiple performance indexes \( J_i \) are defined. Different objectives can often be conflicting, that is the attainment of a better \( J_k \) would lead to a worse \( J_h \), \( \{ h, k \in \{ 1; n \}, h \neq k \} \). Hence, the optimisation in terms of two or more objectives generates a number of possible compromise choices. Therefore, a trade-off decision logic must be introduced to identify an individual solution, and this this is the subject of multi-objective optimisation theory, which will be discussed in section 4.

2.6. Resulting mathematical formulation

In summary, having introduced the dynamics, the path constraints and the boundary conditions and the cost functions, the trajectory optimisation problem can be analytically stated as [85]:

> “Determine the states \( x(t) \in \mathbb{R}^n \), the controls \( u(t) \in \mathbb{R}^m \), the parameters \( p \in \mathbb{R}^q \), the initial time \( t_0 \in \mathbb{R} \) and the final time \( t_f \in \mathbb{R} \) \( \forall t_f > t_0 \), that optimise the performance indexes

\[ J = \Phi[x(t_f), u(t_f), p] + \int_{t_0}^{t_f} \Psi[x(t), u(t), p] dt \]

subject to the dynamic constraints

\[ \dot{x}(t) = f[x(t), u(t), t, p] \]

to the path constraints

\[ x(t) = f[x(t), u(t), t, p] \]

Nonlinear dynamics are natively encompassed, while other cases such as discrete-time dynamics may also be accounted for by adopting adequately relaxed formulations. The optimal control of discrete-time systems is beyond the scope of this article, and for more detail the reader may refer to [82-84].
This is the author uncorrected pre-publication version. This paper does not include the changes arising from the revision, formatting and publishing process. The final paper that should be used (available at http://dx.doi.org/10.1016/j.paerosci.2015.11.006) is:


\[ C_{\text{min}} \leq C(x(t), u(t), t; p) \leq C_{\text{max}} \]

and to the boundary constraints

\[ B_{\text{min}} \leq B(x(t_0), x(t_f), u(t_0), u(t_f); p) \leq B_{\text{max}} . \]

3. Numerical solution methods

Figure 8 outlines a comprehensive tree of techniques that were proposed for the solution of TOP, of which the most commonly adopted will be described in this section. Some of the techniques will not be discussed in detail due to their limited diffusion. For details on temporal finite element methods based on weak Hamiltonian formulation for the solution of optimal control problems the reader may refer to [4, 88]. As briefly mentioned in section 1, for the solution of the TOP two mainstream strategies have been extensively adopted, namely direct methods and indirect methods [4, 61, 80, 81]. In the first class, also defined by the motto “discretise then optimise” the determination of the unknown control function is attempted directly, and this involves the discretisation of the infinite-dimensional TOP into a finite-dimensional Non-Linear Programming (NLP) problem. In the indirect methods, which historically emerged beforehand and are defined by the motto “optimize then discretise”, analytical manipulations based on the calculus of variations are exploited to transform the OCP into a nonlinear Boundary-Value Problem (BVP).

![Diagram of Trajectory Optimisation Solution Techniques](image)

**Fig. 8.** Outline of the techniques proposed for the solution of trajectory optimisation problems.

3.1. Lagrangian relaxation and first order optimality conditions

The Lagrangian relaxation consists in approximating a constrained optimization problem with an unconstrained one. The process involves constructing a new function \( \mathcal{L} \), called Lagrangian, which incorporates the constraints \( c \) multiplied by a vector of additional unknowns, the Lagrange multipliers, \( \lambda \):

\[
\mathcal{L}(x, \lambda) = F(x) - \lambda^T c(x)
\]

The optimisation process is then applied to the Lagrangian. Some additional conditions are introduced to ensure that the solution of the relaxed optimisation problem converges to the solution of the constrained optimisation. The augmented functional in the case of the problem of Bolza can be written as:

\[
J = \int_{t_0}^{t_f} \left[ \psi + \lambda^T (x - f) + \mu^T (x - g) \right] dt + [\phi + \nu^T (x - b)]_{t_f}
\]

where \( \lambda, \mu \) and \( \nu \) are the vectors of Lagrange multipliers respectively for dynamic constraints, path constraints and boundary conditions, with dimensions consistent with the related constraints. In the BVP solution process, some of these Lagrange multipliers are promoted to the rank of adjoints or “co-states” and treated separately. Convergence to the optimality of the
augmented functional is ensured by the first-order optimality conditions. As an example, the set of conditions for a path and boundary constrained problem is:

\[ f_u - \psi_u \lambda = 0 \]  
\[ \int_{t_0}^{t_f} (f_p - \psi_p \lambda) \ dt + (b_p + \phi_p \nu) |_{t_f} = 0 \]  
\[ \lambda - f_x + \psi_x \lambda = 0 \]  
\[ (\lambda + b_x + \phi_x \lambda) |_{t_f} = 0 \]  

### 3.2. Boundary-value problems

BVP are differential problems for which the conditions on the value of the unknown solution are prescribed at the boundary(ies) of the domain. In TOP applications, BVP are usually multidimensional and involve Dirichlet boundary conditions (also called of the first type or fixed). As BVP in the aerospace domain are most commonly encountered when dealing with problems of fluid dynamics, thermodynamics and structural analysis, a large portion of the solution methods were developed for these applications. The reader may refer to [89-92] for a general introduction to the BVP theory, and to [93-95] for more details on the state-of-the-art solution techniques, with particular emphasis on problems arising in the fluid-dynamics domain. If the Lagrange and Mayer terms are quadratic in \( \{x, u, p\} \) and the constraints are linear in \( \{x, u, p\} \) the BVP is linear and some originally proposed techniques to attain a numerical solution of the TOP included the methods of complementary functions, the adjoint variables and the particular solutions [96]. Given that considerable nonlinearities are present in most of the models involved in aerospace TOP, the reduction of nontrivial cases into linear quadratic BVP is typically precluded. Systems involving either Mayer or Lagrange functionals that are non-quadratic in \( \{x, u, p\} \) or constraints that are nonlinear in \( \{x, u, p\} \) can be solved iteratively or exploiting some kind of heuristics. Iterative solution techniques can be either based on gradient methods or on higher order methods [96]. Gradient methods use at most the first derivatives of the function under analysis, while *quasilinearisation* methods use at most the second derivatives of the function under analysis. Both involve the solution of a linear BVP at each iteration. A scalar numerical convergence performance has to be defined to assess the efficiency a specific iterative BVP solution algorithm. The following definitions of convergence performances were proposed [96]:

\[ P = \int_{t_0}^{t_f} N(x - \phi) \ dt + N(\psi) \]  
\[ Q = \int_{t_0}^{t_f} N(\lambda - f_x + \phi_x \lambda) \ dt + \int_{t_0}^{t_f} N(f_u + \phi_u \lambda) \ dt + N(\int_{t_0}^{t_f} (f_p - \phi_p \lambda) \ dt + (g_p + \psi_p \nu) |_{t_f}) + N(\lambda + g_x + \phi_x \lambda) |_{t_f} \]  

where \( N \) is the norm operator. The algorithm should stop when either \( P \) or \( Q \) attain a certain desired tolerance, i.e.:

\[ P \leq \varepsilon_1 \lor Q \leq \varepsilon_2 \]  

with \( \varepsilon_1 \ll \varepsilon_2 \), since the compliance to the constraints (metric \( P \)), shall prevail over optimality (metric \( Q \)) in most practical implementations.

### 3.3. Iterative solution of unconstrained nonlinear programming problems

Similarly to the nonlinear BVP arising in indirect solution approaches, the NLP problems involved in the direct solution of TOP are typically solved by iterative algorithms, for which the basic theory is briefly mentioned hereafter, or through some kind of heuristics. The full NLP theory and the development of computationally efficient NLP solution algorithms are beyond the scope of this article. For more details the reader may refer to [91, 97, 98]. Adopting the n-dimensional Taylor series expansion of \( F(x) \) to the third term we may write:

\[ F(x^{(k+1)}) \equiv F(x^{(k)}) + \nabla^T_F(x^{(k)}) \cdot s^{(k)} + \frac{1}{2} s^{(k)^T} H(F(x^{(k)})) s^{(k)} \]  

Where

\[ s^{(k)} \triangleq x^{(k+1)} - x^{(k)} \]  

\[ \nabla^T_F(x) = \left[ \frac{\partial F(x)}{\partial x_1}, ...\frac{\partial F(x)}{\partial x_n} \right] \]  

\[ H(F(x)) = \begin{bmatrix} \frac{\partial^2 F(x)}{\partial x_1^2} & \cdots & \frac{\partial^2 F(x)}{\partial x_1 \partial x_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial^2 F(x)}{\partial x_n \partial x_1} & \cdots & \frac{\partial^2 F(x)}{\partial x_n^2} \end{bmatrix} \]  

The necessary conditions for a local minimiser, \( \tilde{x} \), are:

\[ \nabla^2 F(\tilde{x}) = 0 \]  

\[ s^T H |_{\tilde{x}} s \geq 0 \]  

Whilst the sufficient conditions for a strong local minimiser are:

\[ \nabla^2 F(\tilde{x}) = 0 \]  

\[ s^T H |_{\tilde{x}} s > 0 \]  

An iterative NLP solution can thus be formulated so that the search direction at step \( k \) based on the n-dimensional Newton method is written:

\[ s^{(k)} = -[H^{-1} \nabla^2 F(x)]_{x^{(k)}} \]  

In a number of aerospace applications, the evaluation of derivatives of the objective function or of the constraints at every step may be computationally cumbersome; therefore various numerical heuristic strategies were studied and implemented in NLP solvers. A number of strategies are based on the recursive update of the Jacobian or Hessian matrices, as [98]:
\[ D^{(k+1)} = D^{(k)} + U(\Delta F, \Delta x) \]  

(26)

For instance, an n-dimensional generalisation of the secant method providing a recursive update of the Jacobian that minimises the Frobenius norm, proposed by Broyden, is:

\[ D^{(k+1)} = D^{(k)} + \frac{(\Delta F - D^{(k)}\Delta x)}{\Delta x^T \Delta x} \]  

(27)

For the Hessian matrix, the Broyden-Fletcher-Goldfarb-Shanno (BFGS) update can be adopted, which is a rank two positive definite secant approximation:

\[ D^{(k+1)} = D^{(k)} + \frac{\Delta g \Delta g^T}{\Delta g^T \Delta x} \Delta x^T D^{(k)} \]  

(28)

Heuristic strategies formulated to recursively determine the inverse of the matrices were less successful. It is also important to observe in this context that the quasi-Newton methods based on recursive updates manifest superlinear instead of quadratic convergence rate.

In addition to recursive updates of matrices, other notable factors for the development of computationally-efficient NLP solvers include the treatment of infeasible, redundant or rank-deficient constraints and of discontinuities, scaling, sparsity and mesh refinements, above others. Considerable factors also involve the integration schemes and the matrix calculus algorithms adopted. For more detail on fast and efficient implementations of iterative NLP solution algorithms the reader may refer to [98].

### 3.4. Indirect methods

As previously mentioned, indirect methods are a family of TOP solution methods in which the solution is attempted by applying the theory of the calculus of variations, a reciprocal method providing a recursive update of the Jacobian that minimises the performance index with respect to the states and controls.

\[ \frac{\partial J}{\partial u} = 0 \]  

(39)

The necessary condition for \( J \) to be minimum is that:

\[ J = \phi(x(t_f), t_f) + \int_{t_0}^{t_f} \{ \mathcal{H}(x, u, t) + \lambda^T f(x, u, t) \} dt \]  

(40)

This leads to the following Hamiltonian function:

\[ \mathcal{H}(x, u, t) = \psi(x, u, t) + \lambda^T f(x, u, t) \]  

(41)

so that:

\[ J = \psi(x(t_f), t_f) - \lambda^T f(x(t_0), t_0) + \int_{t_0}^{t_f} [\mathcal{H}(x, u, t) + \lambda^T f(x, u, t)] dt \]  

(42)

The variation of the performance index with respect to the states and controls is:

\[ \delta J = [\delta J_\phi + \lambda^T \delta x]_{t_f} + \int_{t_0}^{t_f} [\delta \mathcal{H} + \lambda^T \delta f] dt \]  

(43)

The co-state equations are therefore written as:

\[ \mathcal{H}_x \lambda + \lambda^T = 0 \rightarrow \lambda^T = -\mathcal{H}_x \mathcal{L} - \mathcal{H}_u f \]  

(44)

\[ \lambda^T(t_f) = \mathcal{H}_x \phi|_{t_f} \]  

(45)

hence eq. 36 can be rewritten as:

\[ \delta J = [\lambda^T \delta x]_{t_0} + \int_{t_0}^{t_f} \mathcal{H}_u \delta u dt \]  

(46)

the necessary condition for \( J \) to be minimum is that:

\[ \mathcal{H}_u \delta u = 0 \quad t \in [t_0, t_f] \]  

(47)

Whenever the controls are constrained as

\[ C(u, t) \leq 0 \]  

(48)

then for all admissible \( \delta u \), i.e. simultaneously fulfilling all constraints \( C \), we shall have:

\[ \mathcal{H}_u \delta u + \delta \mathcal{H} \geq 0 \]  

(49)

which represents Pontryagin’s Maximum Principle [100]. Although indirect methods for the solution of TOP were
somehow overlooked with the rise of efficient direct solution methods and of the computational capability to quickly solve very large NLP problem, recent research has identified some advantages that a combined approach exploiting both the calculus of variation and NLP may provide. Indirect collocation methods mentioned hereafter are an example of such combined approach, which is actively researched at present.

3.4.1. Indirect shooting

The indirect shooting, also known as simple shooting, is one of the most basic iterative methods to attempt the solution of the BVP. The system of Hamiltonian dynamics is integrated numerically together with an initial guess from \( t_0 \) to \( t_f \). If the obtained terminal conditions differ from the desired boundary conditions by more than a set tolerance, \( \epsilon \), an updated initial guess set is generated and another integration is performed. In practice, the indirect shooting is almost never viable for the solution of aerospace TOP, as the Hamiltonian dynamics have a very bad conditioning, which is further exacerbated as the integration intervals \( t \in [t_0, t_f] \) increase. Furthermore, the integration on long time intervals may require a considerable computation time.

3.4.2. Indirect multiple shooting

As an attempt to resolve the limitations of the indirect shooting, in the indirect multiple shooting the time interval is divided into \( n + 1 \) subintervals. The indirect shooting method is then applied individually to each subinterval, that is, the HBVP are integrated together with the initial guesses on the state and co-state values. After each successful integration, the interface conditions

\[
x(t_i^*) = x(t_i^+) \\
\lambda(t_i^*) = \lambda(t_i^+)
\]

are imposed and a root-finding iteration is introduced to minimise the objective function associated with the discrepancies \( \langle x(t_i^*), \lambda(t_i^*) \rangle - \langle x(t_i^+), \lambda(t_i^+) \rangle \) until a satisfactory threshold is attained. The size of the HBVP is effectively increased due to the inclusion of the interface values of the state and co-state at the boundaries of each interval. By partitioning the integration interval in a sufficiently high number of subintervals, the issues associated with the hyper-sensitivity of the Hamiltonian dynamics can be mitigated and a rapid convergence may be eventually attained.

3.4.3. Indirect collocation

Collocation methods such as the ones introduced in the following section in the context of direct methods can be applied to parametrise the state and co-state of the HBVP associated with the OCP. Methods following this approach are defined indirect collocation methods. Indirect collocation methods were recently considered for the solution of aerospace TOP [80, 101]. These approaches have promising qualities when the application to TOP is considered, though the limitations in terms of flexibility of indirect methods are still challenging. Additionally the analytical derivation performed prior to establish the HBVP can differ considerably based on the TOP formulation, and this notably restricts the applicability and flexibility of the approach, effectively limiting the application of indirect methods to specific problems where their adoption is advantageous, such as in the case of trajectory optimisation in the presence of wind, discussed in section 5.5.

3.5. Direct methods

As previously mentioned, direct methods involve the transcription of the infinite-dimensional problem in a finite-dimensional NLP problem, hence following the approach summarised as “discretise then optimise”. The transcription into finite-dimensional NLP problem can be either performed by introducing a control parameterisation based on arbitrarily chosen analytical functions, as in transcription methods, or by adopting a generalised piecewise approximation of both control and state variables based on a polynomial sequence of arbitrary degree, as in collocation methods. In both cases the transcribed dynamical system is integrated along the time interval \([t_0, t_f]\). The search of the optimal set of discretisation parameters is formulated as a NLP problem, which is solved computationally by exploiting the most efficient numerical optimisation algorithms available. In direct transcription methods, a basis of known linearly independent functions \( q_k(t) \) with unknown coefficients \( a_k \) is adopted as the parameterisation in the general form:

\[
z(t) = \sum_{k=1}^{N} a_k q_k(t)
\]

Considerable research on computationally efficient TOP solution algorithms based on direct transcription was performed by Betts et al. [58, 60, 102-108]. Analogously, remarkable work on computationally efficient TOP solution algorithms based on direct collocation methods was performed by Rao et al. [109-116].

3.5.1. Direct shooting

In the direct shooting and multiple direct shooting, the parameterisation is performed on the controls \( u(t) \) only, the dynamic constraints are integrated with traditional numerical methods such as the Runge-Kutta family, and the Lagrange term in the cost function is approximated by a quadrature approximation. The parameterisation of the control variables is expressed as:

\[
u(t) = \sum_{k=1}^{N} c_k q_k(t)
\]

3.5.2. Multiple direct shooting

In the multiple shooting, the analysed time interval is partitioned into \( n + 1 \) subintervals. The direct shooting method is then applied to each subinterval. Continuity of the state is enforced at the interfaces as in the following expression:

\[
x(t_i^-) = x(t_i^+)
\]
3.5.3. Local collocation methods

By introducing distinct parameterisations on both the control variables and the state variables, it is possible to implicitly integrate the dynamics and then further approximate the integral of the dynamics with a quadrature as:

$$
\int_{t_i}^{t_{i+1}} f(x(s), u(s), s, p) ds \approx \sum_{j=1}^q \beta_j f(x(\tau_j), u(\tau_j), \tau_j, p)
$$  \hspace{1cm} (47)

where \( \tau_j \) are the nodes of the quadrature approximation. The Lagrange term of the cost function \( \psi \) is also approximated with a quadrature integral form. This methodology is at the basis of a family of direct OCP solution methods, named collocation methods.

3.5.4. Global collocation methods

In the global collocation methods, the direct solution of the OCP is attempted by enforcing the evaluation of the state and control vectors in discrete collocation points across the entire problem domain. Pseudospectral Methods (PSM) are considered one of the most computationally effective families of global collocation techniques available for the direct solution of large nonlinear OCP. They are based on a global collocation of orthogonal (spectral) interpolating functions. Due to their considerable success, we review their working principle in more detail. For further reference the reader may refer to [80, 112]. As an initial step, their application involves the introduction of the non-dimensional scaled time \( \tau \in [-1, 1] \) so that:

$$
t = \frac{t_f - t_0}{2} \tau + \frac{t_f + t_0}{2}
$$  \hspace{1cm} (48)

Such change of variable involves that the differential operator is transformed as follows:

$$
\frac{d}{dt} = \frac{2}{t_f - t_0} \frac{d}{d\tau}
$$  \hspace{1cm} (49)

If the final time is not known (i.e. it is either unconstrained or inequality-constrained in the boundary conditions), \( t_f \) will be an additional unknown variable that will have to be determined by the NLP solver. The states and control variables of the OCP are approximated by a set of polynomials of order \( N \), and the problem is thereby discretised in \( N + 1 \) nodes. These interpolation polynomials must be an orthogonal basis in the discretised space. Hence, they have to satisfy the null scalar product property:

$$
P_i(x_k) \cdot P_k(x_l) = 0 \quad \forall \ i \neq j, \forall \ k, l \in \{1,.., N + 1\}
$$  \hspace{1cm} (50)

Although various families of interpolating polynomials can be successfully adopted and comprehensive dissertations may be found in [117, 118], the best implementations of PSM in terms of computational efficiency adopt simple interpolation polynomials in conjunction with a careful selection of the collocation nodes distribution. For such reasons, the basic Lagrange polynomials are frequently adopted for the interpolation of states and controls. A Gaussian quadrature rule ensures the exact results of discrete integral evaluations. Adopting the interpolation polynomials \( P_k(\tau) \) on the \( N + 1 \) nodes \( \tau_k \), the states are approximated as:

$$
\bar{x}_i(\tau) = \sum_{k=1}^N \bar{x}_i(\tau_k) \cdot P_{i,k}(\tau)
$$  \hspace{1cm} (51)

and the controls are approximated as:

$$
\bar{u}_i(\tau) = \sum_{k=1}^N \bar{u}_i(\tau_k) \cdot P_{j,k}(\tau)
$$  \hspace{1cm} (52)

The evaluation of the dynamic constraints (i.e. the state equations) is then performed in the nodes only, leading to a problem of finite dimensions. The dimension of the discrete problem is not the same in all cases, though. Lagrange polynomials of order \( N \) are expressed as:

$$
P_k(\tau) = \prod_{j \neq k}^{N} \frac{\tau - \tau_j}{\tau_k - \tau_j}, \quad \forall \ j \in \{0, N\}
$$  \hspace{1cm} (53)

As an example, a representation of Lagrange polynomials of the 4th order for equally spaced nodes (basic case) is given in figure 9, whereas figure 10 depicts the same polynomials applied to Chebyshev nodes. Chebyshev Pseudospectral Methods (CPM) adopt Chebyshev polynomials of order \( N \). An application of the Chebyshev pseudospectral method to aircraft dynamics is found in [119] and involves the evaluation of the Chebyshev trigonometric polynomials:

$$
P_N(\tau) = \cos(N \cos^{-1} \tau)
$$  \hspace{1cm} (54)

in the \( N + 1 \) nodes:

$$
\tau_k = \cos \frac{k \pi}{N}, \quad k \in \{0,N\}.
$$  \hspace{1cm} (55)

Two recently adopted PSM variants are the Gauss PSM and the Legendre-Gauss-Lobatto (LGL) PSM [120]. Gauss PSM are based on the Gauss-Legendre quadrature, whereas the LGL PSM are based on the LGL quadrature, also simply known as Lobatto quadrature. Gauss PSM are specifically conceived to ensure that the Karush-Kuhn-Tucker (KKT) conditions are identical to the discretised first-order optimality conditions. Legendre polynomials may be calculated by using the Rodrigues formula:

$$
P_{N}(\tau) = \frac{1}{2^N N!} \frac{d^N}{d\tau^N} [(\tau^2 - 1)^N]
$$  \hspace{1cm} (56)

The Legendre-Gauss-Lobatto (LGL) nodes are the \( N + 1 \) zeros of the polynomial:

$$
L_N(\tau) = (1 - \tau^2)P_N(\tau)
$$  \hspace{1cm} (57)

where \( P_N(\tau) \) is the first derivative of the Legendre polynomial of degree \( N \) [121]. In Gauss PSM the dynamic constraints are not collocated at the boundary nodes, whereas in the LGL PSM the evaluation of states and controls is performed also at the boundary nodes, thereby the dimension of the NLP problem is increased by 2 additional nodes.
This is the author uncorrected pre-publication version. This paper does not include the changes arising from the revision, formatting and publishing process. The final paper that should be used (available at http://dx.doi.org/10.1016/j.paerosci.2015.11.006) is:

4. Multi-objective optimality

As mentioned in Section 2, the optimisation in terms of multiple conflicting objectives $J_k = Q_k(p), k \in [1, n]$ leads to large set of solutions that can be considered optimal in some sense that will be expanded in the following section. Therefore, a trade-off selection strategy must be introduced in order to identify a single optimal solution from the large set of compromise solutions, and this is the subject of multi-objective optimisation theory, which will be discussed in this section. Conflicting objectives arise when introducing multiple environmental, economic and operational criteria. Furthermore, as previously mentioned in section 1.4, the implementation of constraints that are either unfeasible or contrasting the attainment of better optimality can hinder the optimisation process or introduce a certain level of ambiguity, which also has to be addressed by adopting suitable multi-objective optimality techniques. This is particularly noteworthy if the operation in real traffic environment is considered, as the arbitrary number of constraints introduced in the online tactical timeframe to resolve traffic conflicts can encumber the optimisation process. For all these instances it is necessary to complement the optimisation algorithm with predefined heuristics, decision logics and Human Machine Interface and Interaction (HMI) formats for that increase the versatility and effectiveness of the trajectory planning algorithm.

A comprehensive and detailed review covering a large number of multi-objective optimisation strategies for engineering applications is presented in [134]. This section resumes the techniques that are particularly suited for the optimisation of aircraft trajectory with respect to multiple operational, economic and environmental criteria, while complying with an arbitrary number of constraints. As we will describe, this decision making process is usually performed either by means of an a priori articulation of preferences (i.e. beforehand), or of an a posteriori articulation of preferences (i.e. afterwards). Some other strategies were proposed and adopted, including progressive articulation of preferences and no articulation of preferences [134]. In the aerospace domain, bi-objective trajectory optimisation, i.e. in terms of fuel costs and flight time costs, has been widely studied and has led to results that have been exploited for some time in the aviation domain. An important advantage is that by limiting the study to two objectives it is possible to introduce a single scalar value, such as the Cost Index (CI) that is implemented in most of the current generation FMS, to account for the trade-off between the two conflicting objectives. In the emerging environmentally sustainable aviation research, it is nonetheless fundamental to account for several different objectives in a flexible multi-model / multi-objective optimisation framework. Pollutant emissions, fuel consumption, perceived noise, weather phenomena, operative costs and contrails formation are examples of the most common objectives currently studied.

3.5.5. Heuristic methods

Although heuristics are involved to a certain extent in most numerical solution algorithms at various levels, in this section we briefly review the solution strategies for aerospace TOP that involved a substantial exploitation of heuristics and metaheuristics for the optimization of aircraft flight trajectories. A number of researchers proposed TOP solution strategies exploiting evolutionary algorithms, as they can natively feature nonlocal search scopes, hence detecting global optimality regions. Evolutionary Algorithms (EA) refer to a class of algorithms emulating the natural evolution processes, which are considered computationally intelligent since their pruning of unsuitable solutions emulates a machine learning process. For mathematical optimisation purposes, a particularly useful evolution process is the natural selection, for which a population of individuals (genotypes) evolve towards the fittest. Cross-overs and mutations are the genetic processes most commonly recreated to generate new off-springs to be added to the population, which may or may not prove fitter. Mutations and crossovers are of nondeterministic nature and allow EA to overcome the strict limitation of the search scope concept, which would otherwise cause convergence to local minimisers if the initial guess is not sufficiently close to the region of global optimality. The most desired quality of EA is therefore that they are natively suitable for the research of the global optima, since the method produces an arbitrary number of non-deterministic initial guesses and the algorithm itself is not restricted on the concept of the search scope. Among other research activities, EA were exploited as part of the research within the Systems for Green Operations Integrated Technology Demonstrator (SGO-ITD) of Clean Sky by the Green Systems for Aircraft Foundation (GSAF) academic cluster [122-133].
4.1. Pareto optimality

A point in the design space \( p^* \in P \) is defined Pareto optimal if and only if there does not exist another point \( p \in P \) such that \( Q(p) \leq Q(p^*) \) and \( Q_i(p) < Q_i(p^*) \) for at least one \( i \). The Pareto front, also called the Pareto frontier, is the set of all the Pareto optimal points \( p^* \). Pareto optimal points are non-dominated, that is, there does not exist another solution that strictly dominates the Pareto optimal solution in terms of any objective. The Pareto front is the multi-objective and multi-dimensional equivalent of the single optimal solution of single objective optimisation. Due to the fact that in many applications a single solution is ultimately sought for even in large complex problems, multi-objective optimisation techniques shall lead to the selection of a single optimal solution that must be Pareto optimal, eventually at least in the weak sense, and thus must belong to the Pareto front.

4.2. A priori articulation of preferences

In the a priori articulation of preferences approach, the user adopts a formulation of the multi-objective optimality that involves either a quantitative or a qualitative combination or prioritisation of the various objectives \( J_k, k \in [1, n] \) and hence leads to the definition of a single combined objective \( f \). A single-objective TOP solution method is employed to attain the combined objective, ultimately leading to an individual optimal solution. This approach is schematically represented in figure 11. Methods falling in the a priori category analyse ways to define the combined objective starting from the various possible preference articulations of the user.

![Fig. 11. Schematic representation of the a priori articulation of preferences.](image)

4.2.1. Weighted global criterion method

A very common choice for the condensation of the various objectives in a single cost function is to assign a weight to each objective function \( Q_i(p) \) or to a functional of the objective and sum them together. By defining \( Q^0 \) the utopia point, that is \( Q_i^0 = \min_{p \in P} Q_i(p) \), we can introduce the following combined weighting:

\[
\tilde{f} = \left( \sum_{i=1}^{n_j} w_i^j [Q_i(p) - Q_i^0]^s \right)^{1/s}
\]

(58)

where \( w_i \neq 0 \forall i \) is the weight assigned by the user to each single objective. Generally \( \sum_{i=1}^{n_j} w_i^j = 1 \), but this is not strictly necessarily. An important subcase is when \( w_i = 1 \forall i, r = 1 \) and \( s = 2 \), so that the weighted global criterion is actually the geometric distance (modulus) of the optimal solutions from the utopia point. Another subcase, which is the simplest and perhaps the most common method to combine the various objectives in a single one is the weighted sum method, for which \( Q^0 = 0 \forall i \), and \( r = s = 1 \), so that

\[
f = \sum_{i=1}^{n_j} w_i Q_i(p) \tag{59}
\]

For all the weighted global criterion methods, an adequate normalisation of the objective functions \( Q \), even if not explicitly mandated, is fundamental to attain the intended balance of importance. This aspect is nontrivial, because the results will noticeably depend on the choice of reasonable reference values for normalisation. The presence of a nonzero utopia point in the generalised formulation partially alleviates this criticality, by replacing absolute magnitudes with relative ones.

4.2.2. Weighted min-max

The weighted min-max, also called weighted Chebychev method, endeavours to optimise the worst performance among the various objectives at any given time, as in:

\[
f = \max_k (w_k [Q_k - Q_k^0]) \tag{60}
\]

The weighted min-max may frequently require more iterations than the weighted global criterion to reach the optimal solution, but at every step the structure of the evaluated \( f = Q(p) \) will be simpler than the weighted global criterion, potentially improving the computational speed when complex nonlinear objective functions are present.

4.2.3. Weighted product

A different kind of weighting was proposed for which the combined performance index is defined as a product of all the objective functions, to the power of the assigned weight \( w_i \), as in:

\[
f = \prod_{i=1}^{n_j} (Q_i(p))^w_i \tag{61}
\]

It is important to note that in order to attain a non-trivial solution, it is strictly necessary that the condition

\[
Q_i(p) \neq 0, \forall i \tag{62}
\]

is enforced at all times. An advantage of this formulation compared to the weighted global criterion is the diminished dependence on the quality of the normalisation performed, so this approach may prove useful when the range of the objective function is unknown or unbounded. A strong limitation of the weighted product, on the other hand, is that when nonlinearities are present in one or more of the objective functions \( Q(p) \), the computational complexity of the optimisation may increase considerably.
4.2.4. Exponential weighted criterion

To overcome the limitation to convex portions of the search domain imposed by weighted global criterion methods, an exponential weighted formulation can be adopted, as in:

\[ J = \sum_{i=1}^{n_j} \left( e^{p_{w_i}} - 1 \right) e^{p_{0}(p)} \] (63)

As in the weighted product, a significant advantage is the diminished dependence on the quality of the normalisation performed, while nonlinearities in the objective functions may still affect the computational performances significantly.

4.2.5. Lexicographic and sequential goal programming methods

Another strategy for a priori articulation of preference is the definition of an order of importance for the various objectives. A single objective optimisation step is then performed for each objective \( f_k \) in the defined order, enforcing that the performance of the new step \( Q(p^{(k)}) \) is equal or better than the previous step \( Q(p^{(k-1)}) \) for all the objectives between 1 and \( (k) \), as in:

\[ Q_j(p^{(k)}) \leq Q_j(p^{(k-1)}), \quad \forall j \in [1; k] \] (64)

Since the original version of the method appears heavily unbalanced towards the first objective, a relaxed version for which in each new step a limited freedom to worsen the performance of the previous step is introduced in order to attain a more balanced optimal solution

\[ Q_j(p^{(k)}) \leq \left( 1 + \frac{\delta_j}{100} \right) Q_j(p^{(k-1)}), \quad \forall j \in [1; k] \] (65)

4.2.6. Physical programming

A notable physical programming approach was proposed, for which the user can introduce intuitive considerations and unstructured information that are used as design metrics to construct the combined objective function. The user may, in particular, define ranges, such as desirable, tolerable, undesirable and unacceptable, to particular parameters, based on qualitative considerations. The user introduced information is then translated in a number of structured numerical objective functions as:

\[ Q_\alpha(p) = \log \left( \frac{1}{dm} \sum_{i=1}^{dm} Q_i(Q_j(p)) \right) \] (66)

In the scope of trajectory optimisation, the most interesting advantage of the physical programming is that the user can directly intervene on the solution region and restrict the part of the Pareto front studied.

4.3. A posteriori articulation of preferences

In the a posteriori articulation of preferences approach, a single optimal solution, belonging to the Pareto front, is chosen after the whole set or a portion of it has been already determined. This approach is schematically represented in figure 12. Methods in this category aim essentially at populating the Pareto front with an even distribution of points, in order to reduce the computational requirements, increase the effectiveness of the selection and so that the user can perform the final choice from a set of sufficiently diversified solutions.

Fig. 12. Schematic representation of the a posteriori articulation of preferences.

4.3.1. Physical programming

Similarly to the a priori implementation, Physical programming can be applied to translate unstructured information supplied by operator into decision criteria. As already discussed, the physical programming approach requires the definition of desirable, tolerable, undesirable and unacceptable ranges of objective functions and constraint values (ranges for the state and control variables are being defined by the problem inequality constraints). An advantage of the a posteriori physical programming approach in a closed-loop implementation is that the method restricts the search domain in addition to the solution region, by discarding the combination of performance indexes that delivered unacceptable solutions. This would prevent computational resources to be wasted in calculating Pareto optimal points which lie outside of the acceptable portions of the Pareto front.

4.3.2. Normal boundary intersection

In the Normal Boundary Intersection (NBI) strategy, the user-supplied weightings \( w \) are modulated in order to obtain an even distribution of Pareto optimal points. The NBI strategy can be formulated as:

\[ \text{Minimise } \lambda, \text{ subject to: } \Phi w + \lambda n = Q(p) - Q^0 \]

where \( \Phi \) is the pay-off matrix made by the column vectors of the objective functions \( Q(p_i) - Q^0 \) at the minimum of the \( i \)-th objective function. The vector \( w \geq 0 \), \( \sum_i^m w_i = 1 \) are parameters to be systematically modified to obtain the complete Pareto front. It must be pointed out that the method does not provide sufficient condition for Pareto optimality, so some generated solutions are actually weakly optimal or suboptimal points.

4.3.3. Normal constraint method

The Normal Constraint (NC) method is a modification of the NBI method for which a tactic to filler suboptimal solution is encompassed. For this method, an arbitrary number of evenly distributed sample points are determined in the utopia hyperplane as a linear combination of the vertices with consistently varied weights. Each sample is then correlated to a Pareto optimal solution through a single objective optimisation process.
5. Optimality criteria and constraints

This section presents a number of optimality criteria and constraints typically implemented as part of the TOP to address the operational, economic and environmental performances of an aircraft in flight and to best represent its realistic operational conditions. Optimality criteria and constraints are introduced in the optimisation problem by means of suitable models. Figure 13 outlines the interdependencies introduced in principle by models adopted in MOTO for environmental sustainability studies.

These models can either consist of one or more mathematical functions, which are either analytically derived from physical/geometrical principles or empirically formulated, or consist of raw numerical data, usually in tabular form. From the theoretical point of view, it is desirable to identify representative functions in analytical form, i.e. being accurate, simple, versatile and universal, though these qualities are frequently conflicting in nature. The identification of functions in analytical form has the advantage of leading to results of high scientific significance and/or of more general applicability, which are the most relevant from the research perspective. Another advantage of adopting functions in analytical form is that the effectiveness of numerical solution strategies can be accurately estimated. In practice, in a number of applications, it is necessary to deal with numerical data in tabular or other forms. Some detail is therefore given to techniques for the exploitation of raw data, and to the necessary pre-arrangements to ensure numerical stability and convergence.

5.1. Flight dynamics

As the concept of flight trajectory is deeply related to the dynamics of the body in aerial motion, flight dynamics are a core component, and are discussed beforehand. The focus of this article is on fixed-wing transport aircraft; hence the models introduced are specifically tailored to this category of flying platforms. A detailed discussion on the derivation of flight dynamics equations from first principles is beyond the scope of the article, and the reader may refer to [135-138] for more information.

5.1.1. Rigid body models

Assuming the aircraft to be a rigid body with a static mass distribution, a rather accurate model of its flight dynamics can be introduced, which complement the equilibrium of forces along the coordinate axes of a suitable Cartesian reference frame located in the centre of mass of the aircraft, named body frame, with the equilibrium of their momentums. This model involves a high number of parameters to define the properties of inertia and of aerodynamic stability and control forces. Adequate experimental and numerical investigations are typically required in order to define the parameters with good precision. For the implementation in TOP, and for other applications including flight simulation and trajectory estimation, flight dynamics are typically transcribed in a set of Differential Algebraic Equations (DAE). An introductive description of such derivation may be found in [135]. The set of DAE and complementary kinematic relations defining the six Degree of Freedom (6DOF) rigid body dynamics of a fixed-wing aircraft are [139]:

\[
\begin{bmatrix}
\dot{u} \\
\dot{\psi} \\
\dot{w}
\end{bmatrix} = \frac{\theta}{W} \begin{bmatrix} F_x \\ F_y \\ F_z \end{bmatrix} + \begin{bmatrix} -r \theta \\ s \phi \psi \\ c \phi \theta \end{bmatrix} + \begin{bmatrix} r v - q w \\ p w - r u \\ q u - p v \end{bmatrix}
\]

where:

\[
\begin{bmatrix}
\dot{p} \\
\dot{q} \\
\dot{r}
\end{bmatrix} = \begin{bmatrix} l_{xx} & 0 & -l_{xz} \\ 0 & l_{yy} & 0 \\ -l_{xz} & 0 & l_{zz} \end{bmatrix}^{-1} \begin{bmatrix} M_{xb} + (l_{yy} - l_{zz})q r + l_{xz}p q \\ M_{yb} + (l_{zz} - l_{zz})p r + l_{xz}q r \\ M_{zb} + (l_{xx} - l_{yy})p q - l_{xz}q r \end{bmatrix}
\]

\[
\begin{bmatrix}
\dot{x} \\
\dot{y} \\
\dot{z}
\end{bmatrix} = \begin{bmatrix} c \theta \psi \\ s \phi \theta \psi \\ l_{yy} \psi \end{bmatrix} \begin{bmatrix} c \phi \theta \psi + s \phi \psi \\ c \phi \theta \psi + s \phi \psi \\ c \phi \theta \psi - s \phi \psi \end{bmatrix} \begin{bmatrix} u \\ \dot{u} \\ \dot{\psi} \end{bmatrix}
\]

\[
\begin{bmatrix}
\delta \phi \\
\delta \psi \\
\delta \theta
\end{bmatrix} = \begin{bmatrix} 1 & s \phi / c \theta & c \phi / c \theta \\ 0 & c \phi & -s \phi \\ 0 & s \phi / c \theta & c \phi / c \theta \end{bmatrix} \begin{bmatrix} \phi \\ \psi \\ \theta \end{bmatrix}
\]

Fig. 13. Layout of the typical model interdependencies in multi-objective trajectory optimisation studies.
Alternative propulsion strategies on medium problem reference frame of the Earth, and were successfully adopted in a number of applications, which are adopted when the analysed timeframes are relatively limited, so that the fuel consumption may be neglected, or when the model can involve either a constant mass or a variable mass. Models belonging to the first category are neglection all aspects associated to the rotational state of the aircraft, as well as the presence of instabilities [140]. Rigid body models are nevertheless fundamental for the study of transition manoeuvres and more so in general for the dynamic stability and control analysis and design of aircraft, and were successfully adopted in a number of trajectory optimisation studies, in combination with a careful selection of path constraints.

5.1.2. Point-mass models

A commonly adopted approach to derive a simplified set of equations of motion for atmospheric flight is based on the approximation of the aircraft as a point-mass object thereby neglecting all aspects associated to its rotational dynamics. The resulting dynamics are characterised by only Three Degrees Of Freedom (3DOF) – i.e. the three spatial coordinates – which is the name adopted to define this family of models. These models are based on Newton’s second law expressed along the coordinate axes of the body frame, and on the expression of the motion of such frame with respect to an inertial reference frame of convenience. All aspects associated with the rotational state of the aircraft are neglected. The model can involve either a constant mass or a variable mass. Models belonging to the first category are adopted when the analysed timeframe is relatively limited, so that the fuel consumption may be neglected, or when no fuel is consumed, such as in the case of sailplanes or total engine failures. As an example, the following set of DAE is associated with a variable mass 3DOF model [135]:

\[
\begin{align*}
\dot{v} &= \frac{g}{W} (T \cos \epsilon - D - W \sin \gamma) \\
\dot{\gamma} &= \frac{g}{v W} \left[ (T \sin \epsilon + L) \cos \mu - W \cos \gamma \right] \\
\dot{x} &= \frac{g}{v W} \left( (T \sin \epsilon + L) \sin \mu \right) \\
\dot{\phi} &= \frac{R_E + z}{v} \left[ v \cos \gamma \sin \chi + v_w \phi \right] \\
\dot{\lambda} &= \frac{v \cos \gamma \cos \chi + v_w \chi}{R_E + z} \\
\dot{\zeta} &= v \sin \gamma + v_w z \\
\dot{m} &= -FF
\end{align*}
\]

where the state vector consists of the following variables:

- \(v\): longitudinal velocity (scalar) [m \ s\(^{-1}\)];
- \(\gamma\): flight path angle (scalar) [rad];
- \(\chi\): track angle (scalar) [rad];
- \(\phi\): geographic latitude [rad];
- \(\lambda\): geographic longitude [rad];
- \(z\): flight altitude [m];
- \(\epsilon\): thrust angle of attack [rad];
- \(m\): aircraft mass [kg];

and the variables forming the control vector are:

- \(T\): thrust force [N];
- \(N\): load factor [ ];
- \(\mu\): bank angle [rad];

Other variables and parameters include:

- \(D\): aerodynamic drag [N];
- \(v_c\): wind velocity, in its three scalar components [m \ s\(^{-1}\)];
- \(g\): gravitational acceleration [m \ s\(^{-2}\)];
- \(R_E\): Earth radius [m];
- \(FF\): fuel flow [kg \ s\(^{-1}\)].

Frequently the modelling is restricted to the flight profile in the vertical plane or in the horizontal plane only.

5.2. Turbofan and turboprop engine models

This section will briefly outline useful models for fuel consumption and thrust with respect to current generation turbofan/turbojet and turboprop engines. Different propulsive technologies are potentially associated with vastly dissimilar energy sources and consumption rates, and this is increasingly evident with the appearance of novel alternative propulsion systems. Notwithstanding, most of the currently operated aircraft rely on air-breathing internal combustion engines of the turbofan or turboprop type, employing hydrocarbon fuels, which are an established technology with efficient and cost-effective production and supply chains. Within the MOTO context, steady state empirical models are commonly employed to reproduce the dependencies of thrust and fuel flow on altitude \(z\), true airspeed \(v_T\), temperature and throttle \(\tau\), which are the only ones directly associated with the aircraft trajectory. The following empirical expressions were adopted in the development of Eurocontrol’s Base of Aircraft Data (BADA), to determine the climb thrust and the fuel flow \(FF\) of a turbofan propelled aircraft, which operationally equates to the maximum thrust \(T_{MAX}\) in all flight phases excluding take-off [141]:

\[
\begin{align*}
T_{MAX} &= C_T \cdot v_T^2 \\
FF &= C_F \cdot v_T^3
\end{align*}
\]
\[ T_{\text{MAX}} = C_{T_1} \left( 1 - \frac{H_p}{C_{T_2}} + C_{T_3} H_p^2 \right) \]  
\[ FF = \max \left[ \tau T_{\text{MAX}} C_{f_1} \left( 1 + \frac{v_{TAS}}{C_{f_2}} \right), C_{f_3} \left( 1 - \frac{H_p}{C_{f_4}} \right) \right] \]

where \( \tau \) is the throttle control, \( H_p \) is the geopotential pressure altitude in feet, \( \Delta T \) is the deviation from the standard atmosphere temperature in kelvin, \( v_{TAS} \) is the true airspeed. \( C_{T_1}, \ldots, C_{T_5}, C_{f_1}, \ldots, C_{f_4} \) are the empirical thrust and fuel flow coefficients, which are also supplied as part of BADA for a considerable number of currently operating aircraft [141]. Similarly, the following empirical expressions were adopted in BADA to determine the maximum thrust and fuel flow of a turboprop propelled aircraft:

\[ T_{\text{MAX}} = C_{T_1} \frac{v_{TAS}}{C_{T_2}} + C_{T_3} \]  
\[ FF = \max \left[ \tau T_{\text{MAX}} C_{f_1} \left( 1 - \frac{v_{TAS}}{C_{f_2}} \cdot \frac{v_{TAS}}{1000} \right), C_{f_3} \left( 1 - \frac{H_p}{C_{f_4}} \right) \right] \]

where again \( C_{T_1}, \ldots, C_{T_5}, C_{f_1}, \ldots, C_{f_4} \) are also supplied as part of BADA for a considerable number of currently operating aircraft [141]. The accuracy of the empirical models and of the coefficients supplied as part of Eurocontrol’s BADA were analysed in [142].

Due to the very competitive specific energies and energy densities, which are furthermore important in the weight/volume sensitive aerospace domain, no significant change in the market-share of hydrocarbon fuels is expected in the near-term future. These aspects, combined with the lengthy and costly development and certification processes to be undertaken by any innovation in the aviation domain, introduce a considerable technological inertia. Given the generational lifespan of aircraft models, it is expected that petroleum-based fuels will still be the largest source of chemical energy for aeronautical propulsion at least for the next two decades. For all these reasons, current generation jet fuels are widely targeted in trajectory optimisation studies for the assessment of aviation environmental impacts in the future.

5.2.1. Pollutant emissions

Emission Index \( (EI) \) specific to each Atmospheric Pollutant \( (AP) \) species were introduced in order to distinguish the dependencies of pollutant emissions from the Fuel Flow \( (FF) \), and are very frequently adopted in TOP studies. The general expression to calculate the total emission of the \( AP \) from its associated Emission Index \( (EI_{AP}) \) expressed in \( [\text{Kg} \cdot \text{AP} / \text{Kg} \cdot \text{hr}] \) is:

\[ AP = \int_{t_0}^{t_f} EI_{AP}(\tau) \cdot FF(\tau, v, z) \, d\tau \quad [\text{Kg}] \]

where \( FF \) is the fuel flow \( [\text{Kg} \cdot \text{s}^{-1}] \).

The ICAO has established an extensive and constantly updated databank for engine emissions based on data collected independently by a number of entities, which proves being a valuable starting point for simplified empirical engine emission models [143]. The fuel-specific \( EI \) are measured at the standard throttle settings defined in ICAO Annex 16 volume 2 [144]. In particular, for an exclusively subsonic engine, the reference throttle settings are take-off (100% of rated engine thrust), climb (85%), approach (30%), and idle (7%).

Eurocontrol’s BADA employs an empirical calculation method for the \( FF \) as a function of the engine thrust \( (\tau) \), of the ambient pressure and of the true airspeed conditions [141]. For the implementation of Eq. 76 in a trajectory optimiser, it is convenient to refer to the differential expression:

\[ \frac{d AP}{dt} = EI_{AP}(\tau) \cdot FF(\tau, v, z) \quad [\text{Kg} / \text{s}] \]

The various aviation-related \( AP \) taken into consideration have different dependencies, which are discussed separately. The most relevant carbon-related exhaust products of fossil fuels are carbon dioxide \( (\text{CO}_2) \), carbon monoxide \( (\text{CO}) \) and unburned hydrocarbons \( (\text{HC}) \). \( \text{CO} \) and \( \text{HC} \) are significantly noxious for both the environment and the living beings, therefore are primarily targeted. Significant amounts of \( \text{CO} \) and \( \text{HC} \) are generated during the incomplete combustion incurring at low throttle settings [145]. An empirical model for \( \text{CO} \) and \( \text{HC} \) emissions \( (EI_{\text{CO/HC}}) \) at mean sea level based on nonlinear fit of turbofan engines experimental data available in the ICAO emissions databank is:

\[ EI_{\text{CO/HC}}(\tau) = c_1 + \exp(-c_2 \tau + c_3) \quad [\text{g/Kg}] \]

where, as a first estimate, the fitting parameters \( c_{1,2,3} \) accounting for the \( \text{CO} \) emissions of 165 currently operated civil turbofan engines from the ICAO emissions database are \( c = (0.556, 10.208, 4.068) \) for \( \text{CO} \) and \( c = (0.083, 13.202, 1.967) \) for \( \text{HC} \) [146]. Figure 14 and 15 represent the experimental data and the empirical models. All the carbon contents of the fuel that are not transformed into \( \text{CO} \) or \( \text{HC} \) are transformed in 1.5 CO2/ton, which has a positive radiative forcing impact and therefore is a major contributor to the greenhouse effect. The reference value is 3.16 tonCO2/tonus-A1 [24].

![Fig. 14. Empirical fit of CO emissions as a function of the throttle for a large number of currently operated turbofan engines](http://dx.doi.org/10.1016/j.paerosci.2015.11.006)
mean sea level based on the throttle setting. The following expression, plotted in Figure 16, is an example of such curve fitting comprehensively accounting for 177 currently operated civil aircraft engines [146]:

\[ E_{\text{NOX}}(\tau) = 7.32 \tau^2 + 17.07 \tau + 3.53 \quad \text{[g/Kg]} \quad (79) \]

![Fig. 15. Empirical fit HC emissions as a function of the throttle for a large number of currently operated turbofan engines [19].](image1)

![Fig. 16. Empirical fit of NOX emissions as a function of the throttle for a large number of currently operated turbofan engines [19].](image2)

In order to obtain an accurate estimate of pollutant emissions at height, a methodology commonly referred to as “Method 2” was developed by Boeing in 1995 [147, 148]. The method proposes an empirical correction to account for installation effects, and subsequently introduces \( E_I \) corrections based on ambient temperature, pressure and relative humidity. The turbofan and turboprop engine models adopted in BADA and presented in section 5.2 implement the altitude dependency as part of the empirical fuel flow models following this philosophy.

For further information on more detailed modelling of aircraft emissions the reader is referred to [34, 131].

5.3. Operational costs

Operational costs of aircraft are a fundamental part of trajectory optimisation. It is necessary in particular to capture the economic dependencies on flight time, which is typically a conflicting criterion with respect to fuel consumption and gaseous emissions. Moreover, some sorts of taxation scheme have dependencies with the flown trajectory, and should also be modelled. By taking into consideration Maintenance Costs (MC), flight Crew Costs (CC), Schedule Costs (SC), Cabin Services (CS) and Fees/Taxes (FT), the Total Costs (TC) of a commercial transport mission can be expressed as:

\[ TC = MC(t) + CC(t) + SC(t) + CS(t) + FT(t, x) \quad (80) \]

Maintenance, crew and ownership costs generally are characterised by a linear dependence on flight time, but crew rotations and shifts introduce step increases when the section is long enough. Schedule costs are a highly nonlinear component, capturing all implications of a suboptimal timetable and fleet exploitation, including unnecessary stopovers in the rotation. Cabin services are also typically nonlinear, for example involving steps when additional on board meals are introduced. Fees and taxes encompass a wide range of different dependencies such as landing/parking fees, carbon taxes etc.

5.4. Atmosphere and weather

The flight of an aircraft is considerably affected by the atmospheric conditions encountered along the flight. Extensive route segments in regions of headwind or crosswind conditions and unforeseen diversions around hazardous weather cells have substantial negative impacts on all environmental and economic performances of a flight, in addition to causing delays that perturb operations and affect passenger comfort. A number of models also include local atmospheric conditions as input. For all these reasons the adoption of an accurate weather forecast model with extensive geographic coverage can contribute substantially to the accuracy of trajectory optimisation studies. A macroscale model such as the Global Forecast System (GFS) of the National Operational Model Archive and Distribution System (NOMADS), published by the National Climatic Data Center (NCDC) of the National Oceanic and Atmospheric Administration (NOAA) periodically provides the global distributions of pressure, temperature, relative humidity, which is an essential input for the contrail model, and winds aloft on extended forecast periods up to 0.25 degrees in resolution. Higher resolution forecast models working at smaller scales provide the forecasts on cloud-base, on the occurrence and dynamics of dangerous clouds, precipitations, fog and haze, clear air turbulence, wind shear and microbursts. Figure 17 depicts the 3-dimensional wind and temperature fields at some of the typical cruise altitudes of a jetliner for a 15-hour advance forecast sampled at 0600 UTC of April 3rd 2015.
The calculation and exploitation of weather conditions and forecasts are actively researched topics, as the influence in the operational efficiency and in the accuracy of TOP results is recognised. For further detail, the reader may refer to [149].

5.5. Trajectory optimisation in the presence of wind

A trajectory optimisation problem of significant interest in the aerospace domain consists in the determination of the optimal routing in the presence of winds, which was originally discussed by Zermelo as early as the 1930s [150, 151], and solicited several decades of active research. The potential environmental and economic benefits offered by an optimal routing in a given wind field are very substantial in most cases and furthermore in case of long-range air transport routes. Notwithstanding, the limited accuracy of medium-to-long term wind forecasts and the increasing air traffic densities that progressively restrict the freedom of lateral and longitudinal routing may compromise, and sometimes revert, the gains. The simplest case consists in determining the optimal routing of an aircraft travelling horizontally at constant cruise airspeed between two known points in a known and constant wind field, ignoring the fuel consumption as well as variations in all other meteorological variables. In addition to loosening the constant altitude constraint, other generalisations involve the adoption of constant or variable non-uniform wind fields (i.e. 2D/3D/4D), more sophisticated modelling of the aircraft dynamics and of the atmosphere, multiply-connected search domains, and more complex cost functions addressing fuel consumption or other objectives in addition to flight time. As this research was largely performed before the recent widespread adoption of direct TOP solution methods, a considerable portion of the proposed numerical solution strategies are based on the calculus of variation. A very comprehensive treatment addressing the theory and solution of the trajectory optimisation problem under space-time varying meteorological conditions was produced by De Jong and the Royal Netherlands Meteorological Institute in 1974 [152]. The study led to the definition of a unified theoretical approach to the problem formulation taking into account both space non-uniformities and time variations of the meteorological conditions and flight performances. Theoretical manipulations led to viable alternative formulations involving gradient equations for the time of transfer and a phase velocity equation for the airspeed. Considerations of the real operational conditions restricting the available choices in flight routing highlighted the suitability of graph optimisation methods for global optimality. Some iterative solution algorithms were proposed, including a special graph algorithm, and their viability in flight planning was numerically demonstrated. Bijlsma discussed the case of optimal aircraft routing in general wind fields and the most promising strategies for the development of a computational solution algorithm [153]. The convergence issues and non-global convergence of iterative solution methods are raised and support the adoption of algorithm based on graph theory. One such method capitalising on the advantages of the calculus of variation and of the graph theory is outlined. The general case with variable cruise speed \( v \) and wind speed \( u \) is introduced by the following dynamic constraints, where the sole control variable, \( \theta \), is the ground track azimuth:

\[
\dot{x}_1 = v(t, x_1, x_2, \theta) \cos \theta + u_1(t, x_1, x_2, \theta) \tag{82}
\]

\[
\dot{x}_2 = v(t, x_1, x_2, \theta) \sin \theta + u_2(t, x_1, x_2, \theta) \tag{83}
\]

Leading to the following Euler-Lagrange equations:
Due to the impracticality of developing trajectory optimisers strictly based on the solution of BVP with the required flexibility to be implemented in avionics and ATM systems, researchers have investigated alternative approaches and heuristics. A method based on mesh discretization that does not require iterative solution is presented in [154]. Another approach based on wind networking was also recently proposed [155, 156]. This approach is especially valuable to increase the trajectory predictability in oceanic airspace and other extents outside primary radar surveillance.

5.6. Noise

The noise generated by various mechanical, thermochemical and fluid dynamic processes within the engine as well as around the aircraft structure is propagated through the atmosphere, which acts as a low pass filter due to thermo-fluid-dynamics and molecular processes [157]. When the receiver is at a sufficient distance, most contributions can be merged into a single point source with non-uniform distribution [157]. By means of a suitable propagation model, and thanks to the adoption of appropriate Demographic Distribution Databases (D3) and Digital Terrain Elevation Databases (DTED), noise footprint on the ground can be determined. Notwithstanding, the high physical complexity associated with the emission and propagation of aircraft noise limits the adoption of accurate models to offline optimisation studies, typically performed to optimise the departure and approach paths of aircraft, which are the portions of the flight in close proximity to the ground. The specific noise problem formulation also depends on the governing regulations, which mandate different output metrics [158-160]. More specifically, due to the interactions between the sound waves emitted by various locations at different frequencies and the aircraft structure at various attitudes and airspeeds, noise emissions are characterised by highly uneven and nonlinear 4D profiles. The vagaries of the atmosphere in terms of unsteady and non-uniform composition and thermodynamic state will further affect the propagation of sound waves and thus alter the noise footprint on the ground. Notwithstanding, some assumptions may enable the integration of simplified noise models in real-time avionics and ATM system implementations for evaluation of an estimated noise footprint of generated trajectories. For instance, the propagation of noise generated by a point source with non-uniform distribution can be formulated adopting a Noise-Power-Distance (NPD) model. The integration of this model with a simplified worst-case emission and with limited resolution and computationally optimised DTED and D3 can enable a quasi-real-time noise model implementation. A number of optimisation case studies specifically targeted the minimisation of perceived noise emissions in the design and redesign of departure and arrival procedures [161-167]. For more detail on the modelling of noise emission and propagation, the reader may refer to [157, 159, 160, 167, 168].

5.7. Condensation trails

Contrails are increasingly addressed in aviation sustainability research due to their recognised radiative forcing potential [29]. Contrails are formed when the combination of local atmospheric temperature and relative humidity are suitable, so that the hot water vapour in the engine exhausts leads to locally attain or exceed liquid saturation conditions, generating droplets that subsequently freeze. The discriminant for such phenomenon is referred to Schmidt-Appleman criterion, and is based on the evaluation of the slope of the exhaust mixing curve (MS):

$$MS = \frac{p \cdot c_p \cdot E_{H_2O}}{\epsilon \left( Q - \frac{\rho v}{\kappa} \right)} \quad \text{[Pa/K]}$$

where \(p\) is the local atmospheric pressure [Pa], \(c_p\) is the specific heat capacity of air [J Kg\(^{-1}\) K\(^{-1}\)], \(E_{H_2O}\) is the water vapour emission index, \(\epsilon\) is the ratio of molar masses of water vapour and air, \(Q\) is the specific fuel combustion heat [J Kg\(^{-1}\)], \(v\) is the true airspeed [m s\(^{-1}\)] and \(\kappa\) is the thrust specific fuel consumption [Kg N\(^{-1}\) s\(^{-1}\)]. The formed contrails are susceptible of convective and thermodynamic processes, which may lead to a rapid dissipation of the formed contrail or to a belated formation. In particular, the theory distinguishes three phases: jet, vortex and dispersion phases, which are governed by different physical processes and are hence modelled individually. An exhaustive description exceeds the scope of this article, therefore the reader is referred to [169-173] for further detail. In relation to the growing environmental and social concerns, during the last decade the research in aircraft trajectory optimisation has increasingly addressed contrail avoidance, and a number of approaches were proposed [174-180]. Among other solutions, a 4D mapping of the lifetime and radiative forcing associated with contrails was proposed by the authors for real-time implementations of contrail models [179, 180]. An altitude-constrained 2-Dimensional plus Time (2D+T) trajectory optimisation case study addressing the conflicting objectives of long-lasting contrails and of fuel consumption is depicted in Figure 19.

Fig. 19. 2D+T MOTO with respect to contrail lifetime and fuel consumption [179].

5.8. Airspace and air traffic models

In addition to reproducing real operational conditions, the trajectory optimisation algorithms shall implement the entire lexicon of ATM route, altitude and airspeed clearances, and translate it in entities that are meaningful for the numerical optimisers. As sections 2 and 3 discussed, these entities are either in the form of boundary conditions, path or dynamics constraints
or cost functions. Therefore, airspace and air traffic modelling both in nominal conditions and in presence of considerable disruptions were studied as part of the trajectory optimisation research. Figure 20 shows the partition of the upper European airspace into sectors at FL310.

Some early works on airspace capacity modelling were published in 1982 [181], and in the 1990s [182, 183] specifically addressing the American National Airspace System (NAS). Although these were based on a number of considerations that are largely still valid, later works evolved substantially from them. The remarkable advances in CNS+A technologies allowed for some enhanced Demand and Capacity Balancing (DCB) strategies. The new systems greatly increase the flexibility of routes and airspace configurations, which are both of great use for DCB. Nevertheless, since at present ATM is still largely based on instructions dispatched verbally by air traffic controllers in individual radio transmissions, workload was identified as being the overarching constraint for airspace capacity in an extensive survey published in 2005 [184]. Recent works such as [185, 186] introduced more complex and accurate models and airspace constructs. Some preliminary Human-In-The-Loop (HITL) evaluations of the novel flexible airspace concepts led to positive acceptance by ATM operators [187], but an identified key factor was the adoption of automated transfer-of-communication and more generally of TBO data-com equipment. Air Traffic Flow Management (ATFM) and conventional airspace management algorithms consider sectors as entities with several entry and exit points that handle traffic entering, exiting and flowing through. This principle has been adopted by many large scale models [185, 188]. Figure 21 depicts two airports O and D where flight f is planned to depart from and arrive at, respectively. In order to arrive to airport D, flight f must pass through either sector \( V_1 \), \( V_2 \) or \( V_3 \), which are characterised by various number of entry and exit points. This permits flight f to access alternative routes in the case one of the original planned sectors becomes unavailable due to severe weather, which is an example of a pre-tactical or tactical ATFM initiative.

![Large-scale network model of airspace sectors](image)

**Fig. 21.** Large-scale network model of airspace sectors [185].

Each sector is represented as a volume of space, which can accommodate a number of traffics depending on the throughput measured at the sector entry/exit points. The traffic either enters from adjacent sectors or originates from an airport within the sector itself. In order to define the network, each sector is connected to other sectors or airports. The traffic flow can be expressed in terms of traffics entering and leaving the sector \( v \) during each time interval \( t \) as:

This is the author uncorrected pre-publication version. This paper does not include the changes arising from the revision, formatting and publishing process. The final paper that should be used (available at http://dx.doi.org/10.1016/j.paerosci.2015.11.006) is:
The number of traffics in the sector, $N_{e,t}$, is equivalent to the number of aircraft previously in the sector $N_{e,t-1}$, and to the number of aircraft entering and leaving the sector at each entry point $e \in E$ and exit point $x \in X$. The variable $k$ here represents any capacity metering element such as airports or sector’s entry/exit points. An overall capacity constraint can be defined on the total number of aircraft simultaneously in the sector: $N_{e,t} \leq N_{MAX}$. When a sector is affected by severe weather, a capacity reduction is dynamically introduced as the weather cell progressively relinquishes portions of the sector volume in terms of altitude and horizontal extents. Capacity constraints for airspace sectors require flights to reroute around most congested sectors. The capacity incurs discrete changes as airspace sectors are consolidated and de-consolidated. An optimised timing of these consolidations and de-consolodations can help minimising the under-capacity instances, by optimally fitting the predicted demand. An overall maximum ACC capacity associated with the available workforce, CNS/ATM infrastructure and airspace geography may pose a hard limit for which sector consolidations and de-consolodations may not be sufficient on their own to prevent under-capacity situations. In these circumstances an ATFM initiative is required, re-distributing the demand on a more uniform profile. An example of the Human-Machine Interface and Interaction (HMI) of current state-of-the-art ATFM Decision Support Systems (DSS) is represented in Figure 22. A single sector is considered and the demand is sampled every 20 minutes. The capacity is depicted in red, and an instance of under-capacity, highlighted in yellow, occurs in correspondence with a temporary capacity reduction.

By introducing in the MOTO cost functions customised penalty terms associated with either the time or the distance travelled within congested airspace sectors, it is possible to numerically promote trajectory planning across less densely trafficked sectors. As an example, figure 23 and 24 depict 2D+T MOTO case studies around congested airspace sectors of the Swiss Area Control Centre (ACC). In parallel with the implementation of 4D-TBO, the ATM research community is tackling the concurrent evolution of the airspace into a dynamically-optimised resource, taking into account traffic, airport and weather updates and forecasts. The long-term evolution of ATFM and airspace management is represented by the Dynamic Airspace Management (DAM) paradigm. While ATFM measures are largely dealing with the modulation of the demand, DAM is conceived to alter the capacity, by overcoming the rigidity of this conventional airspace structure. This can be achieved by modifying the geographic extent of the airspace sectors, morphing their boundaries to accommodate shifting traffic patterns, moving weather or other dynamic factors [19]. The capacity of the neighbouring ATC sectors may be temporarily decreased as a result, but this is a far more flexible concept than consolidating or de-consolidating sectors onto ATC positions. Although regulatory considerations and international agreements restrict the potential DAM measures, the increasing levels of information sharing and interoperability amongst future ATM systems will enable a progressive relaxation of the constraints. A number of different automated partitioning techniques have been evaluated for future DAM DSS implementations [190-192].

![Load Chart Flow](image)

**Fig. 22.** HMI² of THALES TopSky ATFM software [189].

As the dynamics involved in 4D-TBO and DAM are mutually interdependent and both concepts are fundamentally based on real-time optimisation algorithms, significant research is required to address the loop dependence, stability and optimality metrics, promoting rapid convergence to global economic/environmental optimality. With respect to MOTO, in the short-term new airspace models will have to consider dynamically evolving lateral and vertical sector boundaries and continuously varying capacities. In the medium-long term, new combined traffic flow and airspace models integrating 4D-TBO and DAM will be required.
6. CNS+A implementations

New airborne avionics and ground-based ATM DSS are being developed in order to enable 4D-TBO operational concepts and capabilities. This section will outline the current state-of-the-art CNS+A implementations of MOTO, also including some exemplifying case studies from the authors inventory. A number of challenges are known to affect the application of MOTO functionalities in the online contexts, and particularly include:

1. computationally fast and efficient MOTO algorithm implementations, meeting the strict solution time requirements;
2. availability of secure, high-integrity and high-throughput air-to-ground data links for negotiation and validation of the generated 4DT intents;
3. availability of efficient and effective Human Machine Interface and Interactions (HMI²) formats and functions to enable the recalculation, review and customisation of optimised 4DT solutions by the human operators;
4. caching, progressive refinement and adaptation of optimal 4DT solutions within the algorithms to ensure the required reliability levels for dependable adoption of MOTO in automated 4DT planning functionalities;
5. synchronisation of 4DT predictions, constraints and performances between the ground-based systems and the aircraft for mathematical consistency;
6. capability to automatically generate and handle in real-time a considerable number of constraints arising from operation in dense traffic conditions, of which some can be unattainable and shall prompt recomputation of 4DT intents for a number of traffics.

Research is currently addressing these limitations by enhancing the numerical implementations while at the same time outlining the operational and equipage requirements in view of the strict certification requirements.

Modern FMS are the primary airborne DSS providing automated navigation and guidance services. The key components of a
typical FMS architecture are the Flight Management Computer (FMC), Multi Control Display Unit (MCDU), Flight Control Unit (FCU), Electronic Flight Instrument System (EFIS), Multi-Function Display (MFD) and Navigation Display (ND) [146]. Current day FMS provide area navigation (RNAV) services adhering to the Required Navigation Performance (RNP) levels in all flight profiles ensuring accuracy, availability, continuity and integrity by constantly monitoring the performance [193]. A number of integrated navigation and guidance functionalities are adopted in the FMS architecture for meeting the required performance levels [194]. FMS incorporating flight plans have the capability of ensuring the Required Time of Arrival (RTA) and do not have any database size or processing issues [195]. There have been critical evaluations of a number of concepts including an ATM model for generating safe, fuel-efficient, very accurate, and air-ground synchronized 4D-trajectories by using flight segment groundspeed profiles and linking GNSS data to the aircraft FMS with feedback control [195]. A concept of operations of the 4DT aims to negotiation and validation in a CNS+A scenario similar to the “democratic” 4D-TBO paradigm envisaged by Visser [4] is depicted in Figure 25. The Next-Generation FMS (NG-FMS) generates intents consisting of flyable optimal 4DT in order of priority. The provision of multiple 4DT options decreases the negotiation transaction time and the need for remote recalculation of an amended 4DT by the ATM DSS. The ATM DSS system receives multiple preferences of 4DT intents from each manned and unmanned aircraft equipped with NG-FMS (Flow of information I in Fig. 25). Once the optimal conflict free 4DT is identified, the ATM DSS presents the solution to the ATCO for review. Upon operators’ approval, the ATM DSS applies the validated trajectories (Flow of information II) and the airborne DSS sends a confirmation to the ground (Flow of information III) [146].

Fig. 25. Concept of operations of the 4DT intent negotiation and validation in the CNS+A context.

6.2 Targeted operational timeframes

A conceptual definition of the operational timeframes is presented in Figure 26, where the stages specifically targeted in the design of novel CNS+A DSS are encircled. A considerable level of coordination and management is performed before the flight is initiated, i.e. in the offline contexts. These stages are typically associated with less stringent requirements in terms of computational time; therefore a variety of models and optimisation strategies were investigated and substantial research activities are continuing in the subject. Although trajectory optimisation is actively investigated for offline contexts, the specific implementations are beyond the scope of the paper and the reader may refer to [122, 188, 196-204]. On the other hand, the online contexts impose increasingly restrictive requirements as the predicted conflict/hazard approaches.

In the strategic online context, the effect of uncertainties on a longer integration timespan induces a suboptimal estimation, but on the other hand the longer time available for calculation and decision-making enables the adoption of more complex models and algorithms in more extensive search domains. In the tactical timeframe, uncertainties have limited effects, but a safe solution to the problem shall be attained in real-time, thus the involved models, decision logics and the TOP formulation are necessarily simpler and typically based on approximations and linearisations. The emergency phase traditionally involves robust and faultless decision logics for safe avoidance. The delineation between strategic, tactical and emergency are usually set as part of the problem formulation and a distinction based on the type of action to be taken is recently prevailing. In particular, strategic online operations are mostly consisting of tactical ATM actions, initiated by a flow manager before the flight enters the jurisdiction of the controller who will have to implement the measures, whereas tactical online operations can be assumed to occur within the current sector and therefore already responsibility of the jurisdiction controller. As most of the unpredicted disruptions incur while the flight mission is already active, i.e. in the online contexts, substantial expectations are put on the novel automated avionics and ATM DSS that will specifically target these stages. While systems conceived for the emergency context, also called “safety-nets”, will always be based to a large extent on safe avoidance logics, MOTO will play an increasingly higher role in the DSS addressing online strategic and tactical contexts, where significant ATM and ATM duties are exerted.

6.3 System architectures

The architecture of a generic ground-based ATM DSS designed to deploy 4D-TBO in the operational domain is illustrated in Figure 27.
This automated DSS will allow the equipped aircraft to fly user-preferred optimal flight paths without increasing the workload or compromising the situational awareness of flight crews and Air Traffic Control Operators (ATCOs) [146]. This will be attained by relinquishing the repetitive, low-level tasks and restricting their responsibilities to high-level and emergency decisions. Novel ATM DSS will feature enhanced automated navigation services supported by self-separation capabilities to fully exploit the available airspace and airport capacities. In order to cope up with the unpredictable disruptions intrinsically affecting the air traffic flows, novel automated ATM DSS include functionalities for online air traffic flow optimisation considering high-density airspace, re-routing and re-scheduling in real-time. This is accomplished by sharing constraints and restrictions, as well as other relevant aircraft and airport performance and status, and optimisation criteria through enhanced data links in real time. Figure 28 depicts the architecture of NG-FMS, which acts as the airborne DSS and is being developed to integrate 4D-TBO. NG-FMS are the key avionics enablers for generating globally optimal trajectories that fulfil the safety, operational, and environmental requirements in the 4D-TBO context. The key functionalities of the NG-FMS for TBO are MOTO-4D for both flight planning and real-time operations, 4D-TT monitoring, negotiation/validation with the ATM DSS and real-time rerouting and information updating. Specific functions including Sense-and-Avoid (SAA) are incorporated as part of the trajectory predictions module [206, 207]. Integrity management module is used to generate integrity caution (predictive) and warning (reactive) flags based on inputs from different sensors/systems and predefined decision logics. A loss of data leads to re-initialising of the 4DT planning and subsequently the 4DT optimisation process. For instance, the main causes of GNSS signal outage and degradation in flight are identified and modelled to implement integrity thresholds and guidance algorithms in the Avionics-Based Integrity Augmentation (ABIA) system [208, 209].

![Fig. 27. ATM DSS architecture [146].](image1)

![Fig. 28. NG-FMS architecture for remotely piloted aircraft [19].](image2)

6.4. Optimisation of arrival and departure trajectories

Figure 29 and 30 present the results of the optimisation of departure trajectories of a medium-range narrow-body airline transport aircraft implementing the Integrated Noise Model (INM) [160, 210]. The optimisation objectives consisted in minimum noise exposure and fuel consumption. The 70 dB SEL area is one of the key measurable of noise exposure, and was thereby adopted for noise mitigation [159]. Nondominated Sorting Genetic Algorithms (NSGA) were adopted for these offline studies, enabling the identification of global optimality. Both case studies implemented the same engine thrust restrictions. In comparison to the original departure trajectory, the noise-optimised one consists of a less steep initial climb to attain higher airspeed, followed by a steep zoomed segment and subsequently a gradual transition to attain the desired altitude and airspeed. Although these case studies were performed offline with dedicated computing facilities, an implementation of suitably
approximated models in CNS+A DSS for real-time online applications is planned.

6.5. Trajectory optimisation in safety-critical applications

Safety-critical applications of trajectory optimisation algorithms were investigated for airborne emergency DSS, also known as safety-nets. Considerable research on real-time implementations of trajectory optimisation is being performed in connection with the development of SAA systems. As an example, robust parallelised direct shooting solution methods with a posteriori decision logics were implemented for the generation of safe obstacle avoidance trajectories as part of the research on Laser Obstacle Avoidance Systems (LOAS) for manned and unmanned aircraft [211-214]. Figure 31 depicts a number of feasible and unfeasible avoidance trajectories generated by LOAS in real-time in the presence of multiple obstacles. Adopting a multi-phase trajectory optimisation formulation, the selection of the optimal trajectory along the first phase (safe steering) is typically based on minimising a cost function of the form [215]:

\[ J = w_t \cdot t_f + w_d \cdot m(t_f) - w_d \cdot D(t_f) - w_f \cdot \int D(t) \, dt \]  

where \( D(t) \) is the slant distance of the host platform along the avoidance trajectory from the avoidance volume associated with the obstacle, \( t_f = \min\{t_{p>0}\} \) is the time at which the safe avoidance condition is successfully attained, \( m(t) \) is the host platform’s mass and \( \{w_t, w_d, w_f\} \) are the positive weightings attributed to time, distance, integral distance and fuel respectively. In time-critical avoidance applications (i.e., closing-up obstacles with high relative velocities) appropriate higher weightings are used for the time and distance cost elements.

The knowledge in the area of SAA systems and safety-nets for Remotely Piloted Aircraft Systems (RPAS) is evolving rapidly in view of the strict certification requirements. Current research is addressing, in particular, the fusion of cooperative and non-cooperative surveillance data for accurate detection, tracking and safe steering. The reader is referred to the literature in the domain for further detail.

6.6. MOTO algorithm verification and validation

All MOTO algorithms conceived for CNS+A integration require extensive validation by simulation in realistic test scenarios. Simulation cases shall cover all phases of flight and airspace configurations. Particular emphasis is given to the TMA, as trajectories of climbing and descending aircraft usually intersect, requiring frequent tactical ATM interventions, and more generally the constrained airport capacity can lead the arrival traffic to stack, even when disruptions are not present. While concepts such as 4DT point-merge and Continuous Curved Descent Approaches (CCDA) are currently experimented to reduce TMA congestion [216-219], practical implementation aspects are being investigated, especially in the 4DT-TBO context. The Arrival Manager (AMAN) is responsible of the optimised sequencing and spacing of arrival traffic towards a single final approach segment.
The AMAN scenario is the most representative case study of online tactical TMA operations for ATM DSS implementations of MOTO-4D [205]. The ground-based ATM system implementing MOTO-4D algorithms identifies the best arrival sequence among the available options in terms of multiple and conflicting objectives. Longitudinal separation is enforced at the merge-point in the form of path constraints and boundary conditions to ensure sufficient separation upon landing, and to prevent separation infringements in the approach phase itself. Notwithstanding, the ATM DSS implementing MOTO-4D shall be capable to perform point-merge at any metering point. After the initial set of optimal intents has been stored in the ATM DSS, the point-merge sequencing algorithm allocates the available time slots according to suitable optimal scheduling decision-logics. As a reference, assuming a minimum longitudinal separation of 4 nautical miles on the approach path for medium category aircraft approaching at 140 knots, the allocated time slots are characterized by a 90–160 seconds separation depending on the wake-turbulence categories of two consecutive traffics. Example 4D-AMAN simulation results of an ATM DSS implementing MOTO-4D algorithms are depicted in Fig. 32. Waypoints and lines depicted in magenta represent the flyable and concisely-described 4DT consisting of a limited number of fly-by and overfly 4D waypoints, obtained through an operational smoothing algorithm [146]. Fig. 33 depicts the computed 4DT in the AMAN schedule display format. AMAN implementations of MOTO-4D such as the one presented above are actively researched and evolving to address inaccurate trajectory predictions, air-ground trajectory synchronisation issues, unforeseen perturbations and the dynamic handling of multiple altitude/airspeed/path constraints in real-time.

7. Conclusions

Considerable research efforts have addressed the theoretical developments and practical implementations of trajectory optimisation algorithms to improve the economic, operational and environmental performances of air traffic. The progresses in optimal control and nonlinear programming have recently resulted in very efficient numerical solution methods, enabling real-time applications even when complex nonlinear models and multiple constraints are introduced. A number of studies evidenced the significant gains that could be attained by optimising individual manoeuvres, procedures, flight phases or entire flight routes. In several works, the practical implementation of these promising solutions in the civil aviation domain was associated with considerable challenges, as the conventional and largely procedural Air Traffic Management (ATM) operational paradigm would have posed significant restrictions. The adoption of continuously updated 4-Dimensional Trajectory (4DT) descriptors with high navigational accuracy and predictability, in conjunction with data-link based 4DT negotiation and validation functionalities are providing an opportunity for the full exploitation of 4DT optimisation methods into the next generation avionics and ATM systems. Novel avionics and ATM Decision Support Systems (DSS) are currently being researched and developed to introduce the 4DT based Operations paradigms. These systems are designed to deploy advanced air-to-ground trajectory planning, negotiation and validation functionalities to enhance the safety, efficiency and environmental sustainability of air traffic operations. In view of the recognised potential, this article reviewed the current theoretical knowledge on the optimisation of aircraft flight trajectories with emphasis on multiple and often conflicting objectives. Guidelines for the development of efficient Multi-Objective Trajectory Optimization (MOTO) algorithms specifically tailored for novel Communication Navigation and Surveillance/ATM (CNS/ATM) and Avionics (CNS+A) systems were presented and possible implementations were discussed. MOTO algorithms have a clear
potential to enable real-time planning and re-planning of more environmentally efficient and economically viable flight routes by simultaneously addressing the dynamic nature of both weather and air traffic conditions. It is therefore anticipated that a combination of MOTO and Dynamic Airspace Management (DAM) techniques will effectively enable the next step in the CNS+A driven evolutions of the global ATM network.

References

30. L. Sherry, “Improving the accuracy of airport emissions inventories using disparate datasets”, IIE Transactions (Institute of Industrial Engineers), 2014. DOI: 10.1080/0740817X.2014.938845


Conference on Unmanned Aircraft Systems (ICUAS '15), Denver, CO, USA, 2015


216. R. J. de Muynck, T. Bos, A. Kuenz, and S. Toerner, "Implementing time based CDA operations in a medium-high density ATM environment", in proceedings of 28th International Congress of the Aeronautical Sciences (ICAS 2012), Brisbane, Australia, 2012

217. N. Takeichi, T. Hatano, and Y. Fukuda, "Continuous descent trajectory with optimum arrival time controllability", in proceedings of 28th International Congress of the Aeronautical Sciences (ICAS 2012), Brisbane, Australia, 2012
