Universidad de Concepción
Dirección de Postgrado
Facultad de Ciencias Físicas y Matemáticas
Programa de Doctorado en Ciencias Aplicadas
con Mención en Ingeniería Matemática

MARCO TEÓRICO PARA EL PROBLEMA DEL DESPACHO ECONÓMICO/AMBIENTAL

A THEORETICAL FRAMEWORK FOR THE ENVIRONMENTAL/ECONOMIC DISPATCH PROBLEM

Tesis para optar al grado de Doctor en Ciencias
Aplicadas con mención en Ingeniería Matemática

Adrian Carrillo Galvez
concepción-chile
2022

Profesor Guía: Fabián Flores
Departamento de Ingeniería Matemática
Universidad de Concepción, Chile

A Theoretical Framework for the Environmental Economic Dispatch Problem

Adrian Carrillo Galvez

Director de Tesis: Fabián Flores, Universidad de Concepción, Chile.

Director de Programa: Raimund Bürger, Universidad de Concepción, Chile.

Comisión evaluadora

Prof.

Prof.

Prof.

Comisión examinadora

Firma: _____
Prof.

Firma: _____
Prof.

Firma: _____
Prof.

Firma: _____
Prof.

Calificación: _____

Concepción, 2022

Acknowledgments

I would like to especially thank my thesis advisor Fabián Flores Bazán, without whom I am sure it would have been more difficult to walk the path that led me to present this doctoral thesis. Thank you very much for your patience, dedication and trust. I also thank the professors Rodolfo Rodríguez and Raimünd Burger, for all their support in leading the doctoral program; as well as professor Gabriel Gatica, for all the facilities I had, as part of the CI2MA. To the professors who helped me and introduced me, through their courses, in the world of the applied mathematical research, specially professors Manuel Solano and Rodolfo Araya. To the CI2MA and DIM staff, who have always supported me : Lorena Carrasco, Cecilia Leiva, Iván Tobar and Jorge Muñoz.

To my friends and fellow students, who made the experience unforgettable. I reserve a special place of my heart for Yissedt, Yolanda and Néstor, my classmates. Also to all those with whom I coincided in my time as student, especially to those who always had words of encouragement and warm gestures when it was needed the most; to Bryan, Camilo, Cristian, Daniel, Eduardo, Elvis, Felipe, Iván, Joaquín, Mario, Patrick, Paul, Paulo, Rafael, Ramiro, Rodrigo, Sergio and Willian (alphabetically).

To my family, the reason of me doing everything in this life, thank you very much for all you support to my brothers, cousins and uncles. Especially to my wife Ariadna López, for all her patience, love and dedication during this journey that we undertook together, I love you.

I am grateful to the institutions that have financed my studies and research: ANID-Chile by financing my National Doctorate Scholarship through CONICYT-PFCHA/Doctorado Nacional/2017-21170239 and projects FONDECYT 1212004, ACE210010 and BASAL FB210005; to Dirección de Postgrado and Departamento de Ingeniería Matemática of Universidad de Concepción for having financed my stay in scientific events, and to CI2MA for providing the offices and facilities for my Ph.D. studies.

Finally, more than give thanks, I dedicate this thesis to my parents; who are the main reason for which I am here. Thank you for all your sacrifices, for all your love, for all your kindness. To my mother and my father, I love you.

Agradecimientos

Me gustaría agradecer especialmente a mi tutor Fabián Flores Bazán, sin el cual estoy seguro hubiese sido mucho más dificil transitar el camino que me lleva a presentar esta tesis de doctorado. Muchas gracias por la paciencia, dedicación y confianza. Agradezco también a los profesores Rodolfo Rodríguez y Raimünd Burger, por todo su apoyo al frente del programa de doctorado; así como al profesor Gabriel Gatica, por todas las facilidades que tuve como parte del CI2MA. A los profesores que me ayudaron a encaminarme en el mundo de la investigación matemática mediante sus cursos, especialmente los profesores Manuel Solano y Rodolfo Araya. Al personal del CI2MA y del DIM que siempre me brindó su apoyo: Lorena Carrasco, Cecilia Leiva, Iván Tobar y Jorge Muñoz.

A mis amigos y compañeros de doctorado, que hicieron que la experiencia fuese inolvidable. A Yissedt, Yolanda y Néstor, mis compañeros de generación, los llevo en un lugar muy especial. Así también como a todos aquellos con quienes coincidí en mi tiempo de doctorado, especialmente a aquellos que siempre tuvieron palabras de aliento y gestos cálidos cuando más se necesitaban; a Bryan, Camilo, Cristian, Daniel, Eduardo, Elvis, Felipe, Iván, Joaquín, Mario, Patrick, Paul, Paulo, Rafael, Ramiro, Rodrigo, Sergio y Willian (alfabéticamente).

A mi familia, por la que hago todo en esta vida, muchas gracias por todo su apoyo a mis hermanos, primos y tíos. Especialmente a mi esposa Ariadna López, por toda su paciencia, amor y dedicación durante este viaje que emprendimos juntos, te amo.

Agradezco a las instituciones que han financiado mis estudios e investigación: a la Agencia Nacional de Investigación y Desarrollo (ANID) por el financiamiento de la Beca de doctorado nacional a través del Programa de Formación de Capital Humano Avanzado (CONICYT-PFCHA/Doctorado Nacional/2017-21170239) y los proyectos FONDECYT 1212004, ACE210010 y BASAL FB210005; a la dirección de Postgrado de la Universidad de Concepción y al Departamento de Ingeniería Matemática (DIM), por haber financiado mi estadía en eventos científicos, y al Centro de Investigación en Ingeniería Matemática (CI2MA) por proveer las oficinas y facilidades para mi estadía en el doctorado.

Finalmente, más que agradecer, dedico esta tesis a mis padres; quienes son el principal motivo por el cual estoy aquí. Gracias por todos sus sacrificios, por todo su amor, por toda su bondad. A mi madre y mi padre, los amo.

Abstract

In this dissertation several aspects regarding the Combined Environmental Economic Dispatch (CEED) problem were analyzed. The CEED is a real problem concerning different approaches in order to manage the polluting emissions and fuel cost of electrical power networks.
In the Introduction the motivation of the thesis is provided. We described the main components of electrical power networks and its relationship with the different formulations usually employed to handle combined emission/economic objectives. The main contributions and a brief overview of the thesis are also presented.
In Chapter 1 a duality theory approach was proposed for solving the Environmental/Economic Dispatch (EED) problem. The EED is a multiobjective optimization problem, where the polluting emissions and the fuel cost are treated as two conflicting objectives, which are optimized simultaneously subjected to several practical constraints. For the multiobjective problem scalarization, the Weighted Sum Method was used and the associated dual problem was solved by using a quadratic programming algorithm. This strategy was tested on three systems with different number of generators and characteristics. The obtained results were compared with other previously reported, showing some advantages of the proposed approach.
In Chapter 2 another alternative to diminish the polluting emissions released by the generating units was analyzed, the Emission Constrained Economic Dispatch (ECED). This is an optimization problem where the total fuel cost is minimized while treating emissions as a constraint with a prespecified limit. Usually, the fuel cost and emission functions of the generating units must be experimentally derived, introducing then uncertainties in the obtained models. However, these uncertainties are often neglected and the ECED problem is solved considering the coefficients of the functions involved as exact (totally known) values. In this chapter we analyzed the effect of the uncertainties associated to the experimental derivation of input-output curves of thermal power plants. Particularly, when polynomial models are fitted through multiple linear regression, we proposed an approach that, based on the respective prediction intervals, can provide solutions immunized, in some sense, against variability in the coefficients estimates. We tested the proposed approach in a real system from the Chilean electrical power network. For the analyzed system we observed that, when uncertainties are not considered, the deterministic optimal solutions can be environmentally infeasible in some scenarios; whereas solutions obtained through the proposed approach, can significantly diminish the risk of environmental violations. The robustness of the prediction interval-based solutions was obtained with a negligible increase of the total fuel cost in all the cases studied.
In Chapter 3 we analyzed the nonconvex homogeneous optimization problem:

$$\min\{f(x) : \ g(x) = 1, \ x \in C\},$$

where $C \subseteq \mathbb{R}^n$ is a (not necessarily convex) closed cone and f, g are positively homogeneous functions on C with different degree such that $g(x) > 0$ for all $x \in C$, $x \neq 0$. This formulation generalize the Portfolio Problem; which is particularly used to deal with the risks and uncertainties associated to the continuous increase of the share of renewable energy sources, in modern energy portfolios. Once a Lagrangian dual problem is associated, it is provided various characterizations

for the validity of strong duality property. One of them is related to the convexity of $(g, f)(C) + \mathbb{R}_+(0, 1)$, revealing a hidden convexity and a suitable S-lemma. In the case where both functions are of the same degree of homogeneity, a copositive reformulation of the original problem is established. It is also derived a zero-order optimality conditions; KKT (local or global) optimality, giving rise to the notion of L-eigenvalues with applications to symmetric tensors eigenvalues analysis. The case when C is expressed by two quadratic forms is particularly studied. The results can be also applied to a class of quadratic fractional optimization problems with two quadratic constraints, yielding new necessary and sufficient second-order optimality conditions.

Resumen

En este trabajo de tesis se analizan varios aspectos relacionados con el problema del Despacho Ambiental de Cargas. Este es un problema real donde se busca gestionar las emisiones contaminantes y el costo del combustible asociados a la generación de potencia eléctrica.
En la Introducción se proporciona la motivación de la tesis. Describimos los principales componentes de las redes de energía eléctrica y su relación con las diferentes formulaciones empleadas, habitualmente, para gestionar las emisiones contaminantes y los costos. También se presentan las principales contribuciones y una breve reseña de la tesis.
En el Capítulo 1 se propuso un enfoque basado en la teoría de la dualidad para resolver el problema de Despacho Económico Ambiental (EED). El EED es un problema de optimización multiobjetivo, donde las emisiones contaminantes y el costo del combustible se tratan como dos objetivos en conflicto, que se optimizan simultáneamente sujetos a varias restricciones derivadas de las características técnicas del sistema. Para la escalarización del problema multiobjetivo se utilizó el método de sumas ponderadas y el problema dual asociado se resolvió mediante un algoritmo de programación cuadrática. Esta estrategia se probó en tres sistemas con diferentes número de generadores y características. Los resultados obtenidos se compararon con otros reportados previamente, mostrando algunas ventajas del enfoque propuesto.
En el Capítulo 2 se analizó otra alternativa para disminuir las emisiones contaminantes liberadas por las unidades generadoras, el Despacho Económico con Restricciones Ambientales (ECED). Este es un problema de optimización donde el costo total del combustible se minimiza mientras se tratan las emisiones como una restricción con un límite máximo preestablecido. Por lo general, las funciones de costo de combustible y emisión de las unidades generadoras deben obtenerse experimentalmente, por lo que se introducen incertidumbres en los modelos obtenidos. Sin embargo, estas incertidumbres a menudo no son consideradas y el ECED se resuelve considerando los coeficientes de las funciones involucradas como valores exactos (totalmente conocidos). En este capítulo se analizaron las incertidumbres derivadas de la obtención experimental de los modelos de costo y emisiones, así como el efecto en la solución del ECED. En particular, cuando modelos polinomiales se obtienen empleando regresión lineal múltiple, propusimos un enfoque que, basado en los intervalos de predicción correspondientes, puede proporcionar soluciones inmunizadas, en cierto sentido, contra la variabilidad en las estimaciones de los coeficientes. Probamos el enfoque propuesto en un sistema real de la red eléctrica chilena. Para el sistema analizado observamos que, cuando no se consideran las incertidumbres, las soluciones óptimas determinísticas pueden ser ambientalmente inviables en algunos escenarios; mientras que las soluciones obtenidas a través del enfoque propuesto pueden disminuir significativamente el riesgo de violaciones ambientales. La robustez de las soluciones basadas en intervalos de predicción se obtuvo con un incremento insignificante del costo total del combustible en todos los casos estudiados.
En el Capítulo 3 se discutió y analizó el problema de optimización homogéneo no convexo:

$$\min\{f(x) : \ g(x) = 1, \ x \in C\},$$

donde $C \subseteq \mathbb{R}^n$ es un cono cerrado (no necesariamente convexo); f, g son funciones positivamente homogéneas en C, con diferentes grados de homogeneidad y $g(x) > 0$ para toda $x \in C$, $x \neq 0$. Esta

formulación generaliza, en particular, el Problema del Portafolio; herramienta ampliamente utilizada para gestionar las incertidumbres y riesgos asociados a la penetración de fuentes renovables de energía en las matrices energéticas modernas. Varias caracterizaciones son brindadas acerca de la validez de la propiedad de dualidad fuerte respecto al problema dual Lagrangiano asociado. Una de las caracterizaciones está relacionada con la convexidad de $(g, f)(C) + \mathbb{R}_+(0, 1)$, revelando cierta convexidad escondida y un S-lema. Cuando ambas funciones tienen el mismo grado de homogeneidad, se estableció una formulación copositiva del problema original. También son derivadas condiciones de optimalidad de orden cero y condiciones KKT (locales o globales) que originan la noción de L-valores propios con aplicaciones en análisis de tensores simétricos. Se estudió particularmente el problema cuando el cono C está dado por dos formas cuadráticas. Estos resultados fueron aplicados además a una clase de problemas de optimización fraccionaria cuadrática con dos restricciones cuadráticas, obteniéndose nuevas condiciones de optimalidad de segundo orden, tanto necesarias como suficientes. Comparaciones fueron realizadas también respecto a la relajación SDP.

Contents

List of Figures

List of Tables

Introduction

The production, distribution and consume of electricity are global concerns. In recent years and due to the scarcity of energy resources, increasing power generation cost and ever-growing demand for electric energy; optimal economic dispatch has become an extremely important issue in power systems [81]. The economic operation of electric power systems involves two separate steps. The first is the pre-dispatch or selection of equipment to be operated in order to meet the expected loads on the system over some immediate future period of time at minimum total cost [153]. The second step is the on-line economic dispatch which determines, instant to instant, the load to be carried on each unit in operation, in such manner as to minimize the total fuel cost and supplying some technical and functional requirements of the system [164]. This step is known as the Economical Power Dispatch or Economical Dispatch (ED) and it is a constrained optimization problem, with the total load demand as the principally constraint.
In recent years, increasing concerns about global warming and environmental deterioration has drawn more attention on daily optimal operation of electric power systems, since more environmentally friendly generation units need to be operated. Several strategies exist to mitigate both, production and emission of pollutants during the generation of electrical power; but emission dispatch strategies become attractive solutions since it are easy to implement and require less additional cost [50, 143]. The incorporation of environmental concerns to the usual economic dispatch have been treated by authors in different manners. On the Environmental/Economic Dispatch (EED) problem, authors treat the pollution emissions and the fuel cost as two conflicting objectives which are optimized simultaneously subjected to the practical constraints, see the survey in [123]. Whereas in the Emission Constrained Economic Dispatch (ECED) problem, fuel cost is minimized while treating emissions as a constraint with a pre-specified limit [2, 138, 156]. Both problems are analyzed in this thesis work, in order to provide solutions with better performances than the obtained elsewhere. Let us introduce first the corresponding formulations.

Economical Power Dispatch

According to the IEEE [78] (Institute of Electrical and Electronics Engineers), the Economical Power Dispatch is a mathematical optimization problem, that has as objective to minimize the cost of meeting the energy requirements of the system over some appropriate period of time and in a manner consistent with reliable service. The ED is not a modern problem and have been studied for almost a century. In 1962 Noakes and Arismunadar [109] listed references in the field of optimum operation of power systems and related areas. They covered articles published in the United States and Canada since 1919, when presumably power system engineers began to take active interest in this matter. In recent decades the principal subject of research have been the methods of solution of the ED, principally those based on artificial intelligence, although classical and hybrid methods have been recently employed (comprehensive reviews of this subject can be seen in [37, 83, 123, 143, 155]). These investigations also differ by the complexities of the mathematical models considered, differences that are related with the constraints and the specific shape of the cost function selected.

The classical formulation of the ED problem is the following:

$$\begin{aligned} & \min \ F(p) \\ st \quad & 1^\top p = P_D + P_L, \\ & 0 \leq p_{\min} \leq p \leq p_{\max}, \end{aligned}$$

and will be detailed next.

Cost function

The total fuel cost is directly related with the specific type of power generating plant. If N generating units are to be dispatched, the total fuel cost is:

$$F(p) \doteq \sum_{i=1}^{N} f_i(p_i).$$

where each $f_i : \mathbb{R} \to \mathbb{R}$ is the fuel cost function of the corresponding generating unit. The most used types are those based on thermal generators which burn fossil fuel (natural gas, coal, oil). For these generating units, there exist several models generally used to describe its fuel cost. Most authors group these models on smooths or non-smooths [140].

Smooth Models

In these kind of models, mainly, the fuel cost curve is determined by polynomial functions of the generated power [126, 128, 140]:

$$f(p) = a_0 + \sum_{j=1}^{M} a_j p^j. \tag{1}$$

By simplicity, usually the Generation Companies (GENCOs) consider fuel cost as a linear function of the generated power, or as a convex quadratic function of the power, i.e., $M \in \{1, 2\}$ and $a_2 \geq 0$ in (1).

Non Smooth Models

However, in real operation of the generating units, several aspects can affect the smoothness of the fuel cost functions. Nowadays and partially motivated by the increase of the computationally capacities, the investigators have been representing fuel cost functions by complex models that take in count, for example, the valve-point effect. This phenomenon occurs in multi-valve steam turbines, producing a ripple-like heat rate curve and motivates the addition of a sinusoidal term [7]:

$$f(p) = a_0 + \sum_{j=1}^{M} a_j p^j + \left| b \sin\left(c(p_{\min} - p)\right) \right|.$$

According to [7], this kind of model increase the accuracy of the results, but evidently adds more burden on the calculation process, specially by moving us away of the convexity assumptions. In Figure 1 we can observe the shape of the fuel cost function when steam regulator valves are considered.

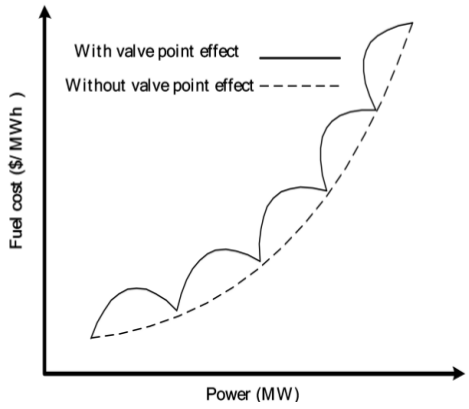


Figure 1: Fuel cost curve with valve-point effect [1].

Also, the generating units can have what it is known as prohibited operation zones. In this case, discontinuities are introduced in the formulation, since now there exist zones where the generating unit can not be operated, mainly due to technical aspects, such as shaft vibration or steam valves characteristics. In Figure 2, a fuel cost curve for a generating unit with two prohibited operating zones is shown.

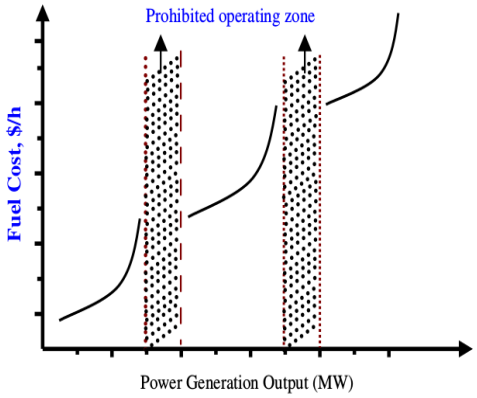


Figure 2: Fuel cost curve with prohibited operation zones [94].

Other cost elements can be incorporated to the objective function, depending on the decision-maker (DM) needs, for example:

- Cost due over-generation and under-generation of power [80, 102]
- Maintenance and operation cost [91, 102]
- Start-up cost (cold or hot) [102]
- Ramp cost [76]

- Spinning reserve cost [76]
- Transmission cost [65]

Problem Constraints

Electrical power networks are very complex systems. Topology of such networks can be very different, but in general exist several participants that must be coordinated in addition to the GENCOs: transmission companies (TRANSCOs) and distribution companies (DISTCOs) [154]. In Figure 3 we can observe a simple but illustrative power network. The electric energy produced by the GENCOs is transmitted by the TRANCOs, which own and operate transmission lines. The DISTCOs own and operate distribution lines, which is the final stage in the delivery of electric power; it carries electricity from the transmission system to individual consumers. This is made through distribution substations that are connected to the transmission system and have as objective to diminish the transmission voltage to medium voltages required by the final consumers. This complexity is incorporated to the ED problem via the considered constraints, that represent the technical and operational boundaries of the whole system [3].

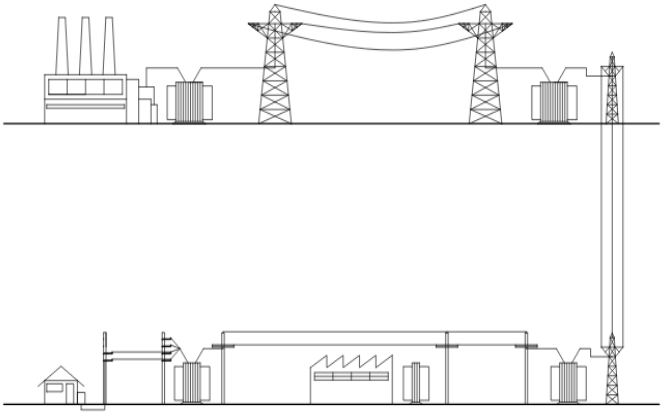


Figure 3: Tipical electric network.

1. Generation Constraints

Power Balance

A particularity of the power network systems is that electrical energy can not be stored in big quantities. Then, the total power generation must be equal to the total load demand; additionally, power losses on the transmission line may be modeled as well:

$$1^\top p - P_D - P_L = 0,$$

where $p = (p_1, p_2, \ldots, p_N)^\top$ is the vector of the power output of each generating unit; P_D is the power demand and P_L is the total transmission losses. The real power loss is obtained from

calculation of the AC load flow problem, which has equality constraints on real and reactive power at each bus as follows [103]:

$$P_{G_i} = P_{D_i} + V_i \sum_{j=1}^{NB} V_j [G_{ij} \cos(\delta_i - \delta_j) - B_{ij} \sin(\delta_i - \delta_j)],$$

$$Q_{G_i} = Q_{D_i} + V_i \sum_{j=1}^{NB} V_j [G_{ij} \sin(\delta_i - \delta_j) + B_{ij} \cos(\delta_i - \delta_j)],$$

where NB is the number of buses; P_{G_i} y Q_{G_i} are the real and reactive power generated at the ith bus respectively; P_{D_i} y Q_{D_i} are the ith bus load real and reactive power respectively; G_{ij} and B_{ij} are the transfer conductance and susceptance between bus i and j respectively; V_i and V_j are the voltage magnitudes at bus i and j respectively; δ_i and δ_j are the voltage angles at bus i and j respectively. There are several methods for solving the resulting nonlinear system, among the most popular is the Newton-Raphson method. The load flow solution gives all the bus voltage magnitudes and angles that can be used to calculate the transmission losses as follows:

$$P_L = \sum_{k=1}^{NL} g_k [V_i^2 + V_j^2 - 2V_i V_j \cos(\delta_i - \delta_j)],$$

where NL is the number of transmission lines and g_k is the conductance of the kth line that connects bus i to bus j. Other authors determine the transmission loss by Kron's loss formula [101]:

$$P_L = \sum_{i=1}^{N} \sum_{j=1}^{N} p_i B_{ij} p_j + \sum_{i=1}^{N} B_{0_i} p_i + B_{00},$$

where p_i, p_j are the power of generators i and j respectively and B, B_0 and B_{00} are loss coefficients.

Generation limits
Each generator should produce power within lower and upper limits, limits that are set according to specific technical parameters:

$$p_{\min} \leq p \leq p_{\max}.$$

Prohibitive operation zones
As previously discussed, in some generating units the whole operating range is not always available for load allocation. These prohibited operation zones introduce discontinuities in the cost function that are modeled as [112]:

$$\begin{aligned} p_i^{\min} \leq \ & p_i \ \leq \ p_{i,1}^L, \\ p_{i,k-1}^U \leq \ & p_i \ \leq \ p_{i,k}^L, \ k = 2, ..., n_i, \\ p_{i,n_i}^U \leq \ & p_i \ \leq \ p_i^{\max}, \end{aligned}$$

where $p_i^{\min}$ and $p_i^{\max}$ are the ith unit minimum and maximum generation limits respectively; n_i is the prohibited zone number; k is the index of prohibited zones of an unit; $p_{i,k}^{L,U}$ are the lower and

upper bounds of the kth prohibited zone of unit i respectively.

2. Security constraints

Line flow
The line flow constraint is used to avoid undesired line loadings due to power distribution. Thus, transmission line loading S_l is restricted by its upper limit as [113]:

$$S_{l_i} \leq S_{l_i}^{\max}, \quad i = 1, 2, ..., N_l,$$

where N_l is the number of lines of the system.

Apparent power flow
Sometimes it is also necessary to limit the apparent power flow (MVA) [146]:

$$|LF_{f_i}| \leq LF_{f_i}^{\max}, \quad i = 1, 2, ..., N_T,$$

where LF_{f_i} is the apparent power flow in line f_i; $LF_{f_i}^{\max}$ is the maximum limit for the apparent power flow at line f_i and N_T is the total number of transmission lines.

System Spinning Reserve Constraint
To achieve a reliable and secure operation in power generation systems, it is usual to considerate a Spinning Reserve. This reservation of power is saved for if encounter any unanticipated operation condition, as unexpected outage of generating units or sudden increase in demand. In [163] this constraint is modeled as:

$$\sum_{i=1}^{N} SP_{G_i} \geq SP_R, \qquad SP_{G_i} = \begin{cases} 0, & \forall i \in \Omega \\ P_{G_{\max_i}} - P_{G_i}, & \text{other} \end{cases}$$

where Ω is the set of generators with prohibited operating zones, SP_{G_i} is the spinning reserve contribution of generator i; SP_R is the system spinning reserve requirement; N is the total number of generating units and P_{G_i} and $P_{G_{\max_i}}$ are the active power and its maximum limit for the ith generator, respectively.

Combined Economic Emission Dispatch

On another hand, the environmental change sweeping the world and its visible effects has motivated the developing of more environmental friendly power production systems. According to [118], since 70's the study of the contribution of the power industry to deteriorating its surrounding has become a major social task.
Power generation with fossil fuels comes with significant costs to the environment and human health [68]. Combustion releases emissions of sulfur dioxide (SO_2), nitrogen oxides (NO_x), particulate matter (PM), carbon monoxide (CO), volatile organic compounds (VOCs) and various trace metals like mercury, into the air through stacks that can disperse this pollution over large areas. Environmental effects of the above mentioned pollutants include impaired visibility, damage to materials, damage to vegetation, deposition as acid rain, ozone formation and contribution to the greenhouse effect [100]. On another hand, in Table 1, major air pollutants and their associated health hazards are presented.

Table 1: Pollutants and their hazards [159].

Name of Pollutant	Health Impacts
Respirable PM	Respiratory illness, including chronic bronchitis and asthma; heart diseases
SO_2	Heart diseases; respiratory problems including pulmonary emphysema, cancer, eye burning, headache
NO_2	Lung irritation, viral infection, airway resistance, chest tightness
Suspended PM	Pneumoconiosis, restrictive lung diseases, asthma, cancer
Ozone	Impaired lung function, chest pain, coughing, irritation of eyes and nose
CO	Cherry lips, unconsciousness, death by asphyxiation

Therefore, emission problems corresponding to the fossil fuels-based power plants cannot be neglected. According to [95], many strategies like installation of air filters, pollutant-cleaning equipment, switching to low emission fuels, replacement of aged equipment and generating units and integration of renewable energy sources, have taken into consideration in order to minimize emissions. But all of them need considerable amount of capital and huge replacement/modification or upgrade strategy that can be considered as a long-term solutions. Thus, emission dispatch strategies (where environmental concerns are now considered in the traditional ED formulation) become attractive solutions, since are easy to implement and require less additional cost [50, 143]. In order to represent the emissions, several models are generally used. Mostly, CO_2 is considered as a linear function of the power output and SO_x as proportional to the thermal unit's fuel consumption, and the corresponding emission function has the same form in (1). Whereas NO_x emissions are more difficult to evaluate because they come from two different sources, nitrogen in the air and in the fuel, and their production is related to several factors including boiler temperature and air content. Thus, the output curve of NO_x can be characterized by an equation consisting of quadratic or linear functions plus additional exponential therms [103, 133]. Then, according to the polluting emission to assess, we must considerate the total emission as:

$$\begin{aligned} E(p) &\doteq \sum_{i=1}^{N}\sum_{j=1}^{M} \alpha_{0i} + \alpha_{ji} p_i^j, \text{ or} \\ E(p) &\doteq \sum_{i=1}^{N}\sum_{j=1}^{M} \alpha_{0i} + \alpha_{ji} p_i^j + \xi_i e^{\lambda_i p_i}, \end{aligned}$$

where α_{ji}, ξ_i and λ_i are the emission coefficients of the ith generating unit.

In Chapter 1 we analyze the Environmental/Economic Dispatch problem:

$$\begin{aligned} \min \ & (F(p), E(p)) \\ st \quad & 1^\top p = P_D, \\ & p_{\min} \leq p \leq p_{\max}, \end{aligned} \tag{2}$$

where, $F(p) \doteq \sum_{i=1}^{N}\sum_{j=1}^{M} \hat{a}_{0i} + \hat{a}_{ji} p_i^j$ is the total fuel cost; $E(p) \doteq \sum_{i=1}^{N}\sum_{j=1}^{M} \hat{\alpha}_{0i} + \hat{\alpha}_{ji} p_i^j$ is the total emission and system is considered as losseless. An approach based on duality theory is developed and results are tested on several systems. By taking advantage of the smoothness and convexity of

the functions involved, the proposed approach provide us solutions that dominate those reported by other authors and obtained by metaheuristic algorithms instead. This chapter is mainly based on paper [30]:

- Adrian CARRILLO-GALVEZ, Fabián FLORES-BAZÁN, Enrique LÓPEZ: A duality theory approach to the environmental/economic dispatch problem. Electric Power Systems Research, vol 184, Art. Num. 106285, (2020).

We must point out that the above cited paper is the result of previous contributions published on the following conferences proceedings [31, 32]:

- Adrian CARRILLO-GALVEZ, Fabián FLORES-BAZÁN, Enrique LÓPEZ: An Analytical Approach to the Environmental/Economic Dispatch Problem. 2019 IEEE CHILEAN Conference on Electrical, Electronics Engineering, Information and Communication Technologies (CHILECON), (2019), pp. 1-5, doi: 10.1109/CHILECON47746.2019.8987495.
- Adrian CARRILLO-GALVEZ, Fabián FLORES-BAZÁN, Enrique LÓPEZ: On the solution of the Environmental/Economic Dispatch problem using Lagrangian duality. 2020 IEEE International Conference on Industrial Technology (ICIT), (2020), pp. 619-623, doi: 10.1109/ICIT45562.2020.9067261.

However, by analyzing the existing literature we became aware that there exist a previous problem that, although extremely closely related to the emission dispatch, is consistently ignored. We point out that when solving (2), all the coefficients of the emission and fuel cost functions are usually considered as fixed (exactly known) values. This deterministic approach can lead to significantly errors in the proposed optimal generation schedule. Note that, in real operation of electric power networks, GENCOs must to establish the fuel cost and emission functions of the generating units. This can be done by several methods: performance testing, determination from operating records or use of manufacturer's guarantee data adjusted to actual operating conditions [78]. But, high cost of performance testing and possible incorrect representation of the shape of the input-output curves provided by manufacturer's data, makes the determination from operating records a suitable alternative; even more if we consider that, nowadays, operating data is readily available in many generation companies. This strategy clearly introduces uncertainties on the experimentally derived models, uncertainties that can lead to obtain dominated solutions (if the EED problem is considered in its deterministic formulation (2)); or to infeasible environmental solutions (if the ECED deterministic formulation is considered). In Chapter 2 we analyze the effect of coefficient estimates in the ECED problem. We compare optimal generation schedules obtained by solving the deterministic or the proposed approaches. The proposed approaches are based on statistical intervals, particularly on the corresponding prediction intervals involved when multiple linear regression is used to fit the models. In fact, since we have sum of response variables, a prediction interval for this sum must be obtained. The proposed approach is also tested on a real system from the Chilean electrical power network, consisting on 4 coal-based generating units. In this case, we can observe the negative effects associated to ignoring uncertainties on models coefficients. This can produce health affectations, particularly during the April-August period, when ventilation and dispersion of air pollutants is critically deficient. The results of this chapter are mainly based on document:

- Adrian CARRILLO-GALVEZ, Fabián FLORES-BAZÁN, Enrique LÓPEZ: Effect of Models Uncertainties on the Emission Constrained Economic Dispatch. A Prediction Interval-Based Approach. **Submitted to publication.**

The results obtained in the second part of this thesis are more general than the obtained in the first part. Although was developed mainly motivated by a specific problem, as we can see in Chapter 3 results obtained were (are) broadly applicable and transcend the merely management of emissions and cost for electrical power networks. Consider that, as was pointed out in [116], modern energy systems and markets are overwhelmed by various sources of uncertainties, due to the renewable energy (RE) revolution in recent years. The value of investments in RE technologies has increased rapidly over the last decade as a result of political pressures to reduce carbon dioxide emissions and the policy incentives to increase the share of RE in the energy mix. As the number of RE investments increases, so does the need to measure the associated risks throughout planning, constructing and operating these technologies. Energy developers, investors and policy makers face a future that implicitly involves technological, financial and political risks and uncertainties. Although, RE technologies potentially have a lower risk profile than conventional energy sources because they are disconnected from fossil fuel prices; they still entail considerable technological, financial and regulatory risk exposure, depending on the technology, country and regulatory regime. Fluctuation of cost components of power generation units, volatile crude oil prices, electricity price and carbon costing in the context of the global climate change mitigation strategy, are examples of uncertainty components encountered by energy developers, investors and policy makers in the energy sector [79]. On another hand, as pointed out in [19], a major drawback for renewable energies in a competitive electricity market is their high price tag, which reduces their competitiveness. Despite the above, RE is still advantageous in terms of cost if the monetary costs of embedded risks related to conventional power are factored into the price of conventional power; and a very suitable strategy to diminish polluting emissions. In this context, a useful tool for effectively manage risk and uncertainty in sustainable energy system planning, is provided by mean-variance portfolio (MVP) analysis, see for example [79, 116] or review in [117].
In general the (classical mean-variance) portfolio optimization problem [97, 98] can be stated as an extension of the standard quadratic optimization problem (StQOP):

$$\mu_q \doteq \min\Big\{\frac{1}{2}x^\top Ax:\ e^\top x = 1,\ x \in C\Big\},$$

where A is a real symmetric matrix of order n, $C \subseteq \mathbb{R}^n$ is a pointed, closed, convex cone having non-empty interior, and $e \in \operatorname{int} C^*$. Here, C^* is the non-negative polar cone of C. In Chapter 3 we not only analyze the MVP problem, but a broad class of minimization problems where the objective and the single constraint functions are positively homogeneous of possibly different degree, and, in addition, a geometric constraint set being a (not necessarily polyhedral) convex cone is considered, i.e., we discuss and analyze the nonconvex homogeneous optimization problem:

$$\min\{f(x):\ g(x) = 1,\ x \in C\},$$

where $C \subseteq \mathbb{R}^n$ is a (not necessarily convex) closed cone and f, g are positively homogeneous functions on C with different degree such that $g(x) > 0$ for all $x \in C$, $x \neq 0$. Once a Lagrangian dual problem is associated, it is provided various characterizations for the validity of strong duality property. One of them is related to the convexity of $(g, f)(C) + \mathbb{R}_+(0, 1)$, revealing a hidden convexity

and a suitable S-lemma. In the case where both functions are of the same degree of homogeneity, a copositive reformulation of the original problem is established. It is also derived zero-order optimality conditions; KKT (local or global) optimality, giving rise to the notion of L-eigenvalues with applications to symmetric tensors eigenvalues analysis. The case when C is expressed by two quadratic forms is particularly studied. The results are then applied to a class of quadratic fractional optimization problems with two quadratic constraints, yielding new necessary and sufficient second-order optimality conditions. Contrary to the SDP relaxation approach (yielding tightness, which in turns implies the fulfillment of that standard strong duality), the present approach goes beyond. The results of this chapter are mainly based on [29]:

- Adrian CARRILLO-GALVEZ, Fabián FLORES-BAZÁN: Non convex homogeneous optimization: a general framework and applications to quadratic fractional programming and tensor eigenvalues analysis. **Submitted to publication**. (Preprint available at: https://www.ing-mat.udec.cl).

Introducción

La producción, distribución y consumo de electricidad son preocupaciones globales. En los últimos años el despacho económico de electricidad se ha convertido en un tema extremadamente importante en los sistemas de energía, debido, entre otras cosas, a la escasez de recursos energéticos, el aumento del costo de generación de energía y la creciente demanda de energía eléctrica [81]. La operación económica de los sistemas de energía eléctrica implica dos pasos separados. El primero es la selección de los equipos que operarán para cumplir con las cargas esperadas en el sistema durante un período de tiempo futuro y a un costo total mínimo [153]. El segundo paso es el despacho económico on-line que determina, instante a instante, la carga a generar en cada unidad en operación, de tal manera que se minimice el costo total de combustible y se cumplan con los requerimientos técnicos y funcionales del sistema [164]. Este paso se conoce como Despacho Económico de Energía o Despacho Económico (ED) y es un problema de optimización con restricciones, con el balance de carga como la restricción principal.
En los últimos años, la creciente preocupación por el calentamiento global y el deterioro ambiental ha llamado más la atención sobre el funcionamiento diario óptimo de los sistemas de energía eléctrica, ya que es necesario operar unidades de generación más respetuosas con el medio ambiente. Existen varias estrategias para mitigar tanto la producción como la emisión de contaminantes durante la generación de energía eléctrica; pero las estrategias de despacho de emisiones constituyen soluciones atractivas ya que son fáciles de implementar y requieren menos costos adicionales [50,143]. La incorporación de las preocupaciones medioambientales al despacho económico ha sido tratada por los autores de distintas formas. En el problema del Despacho Económico Ambiental (EED), los autores tratan las emisiones contaminantes y el costo del combustible como dos objetivos en conflicto que se optimizan simultáneamente y están sujetos a restricciones prácticas [123]. Mientras que en el problema del Despacho Económico con Restricciones Ambientales (ECED), el costo del combustible se minimiza mientras se tratan las emisiones como una restricción con un límite máximo preestablecido [2, 138, 156]. Ambos problemas se analizan en este trabajo de tesis, con el fin de aportar soluciones con mejores desempeños que las obtenidas en otras investigaciones. Introduzcamos primero las formulaciones correspondientes.

Despacho Económico

Según el IEEE [78] (Institute of Electrical and Electronics Engineers), el Despacho Económico de Energía (ED) es un problema de optimización matemática, que tiene como objetivo minimizar el costo de cumplir con los requerimientos energéticos del sistema eléctrico durante un período de tiempo apropiado y de manera consistente con un servicio confiable. El ED no es un problema moderno y se ha estudiado durante casi un siglo. En 1962, Noakes y Arismunadar [109] enumeraron referencias en el área del funcionamiento óptimo de los sistemas de energía y áreas relacionadas. Los autores recopilaron artículos publicados en los Estados Unidos y Canadá desde 1919, cuando presumiblemente los ingenieros de sistemas de energía comenzaron a interesarse activamente en este asunto.
En las últimas décadas el principal tema de investigación ha sido los métodos de solución del ED,

principalmente los basados en inteligencia artificial, aunque recientemente se han empleado métodos clásicos e híbridos (se pueden ver las revisiones exhaustivas de este tema realizadas en [37, 83, 123, 143, 155]). Estas investigaciones también se diferencian por los distintos modelos matemáticos considerados, diferencias que están relacionadas principalmente con las restricciones y la forma específica de la función de costo seleccionada.
La formulación clásica del ED es:

$$\begin{aligned} & \min \ F(p) \\ st \quad & 1^\top p = P_D + P_L, \\ & 0 \leq p_{\min} \leq p \leq p_{\max}, \end{aligned} \tag{3}$$

y será detallada a continuación.

Función costo de combustible

El costo total del combustible está directamente relacionado con el tipo específico de planta generadora de energía. Si N unidades generadoras son despachadas, el costo total del combustible es:

$$F(p) \doteq \sum_{i=1}^{N} f_i(p_i),$$

donde cada $f_i : \mathbb{R} \to \mathbb{R}$ es la función de costo de combustible de la unidad generadora correspondiente. Los modelos más utilizados corresponden a unidades generadoras que consumen combustibles fósiles (gas natural, carbón, petróleo). Para estas unidades generadoras, existen varios modelos que generalmente se utilizan para describir su costo de combustible. La mayoría de los autores agrupan estos modelos en suaves o no suaves [140].

Funciones costo de combustible suaves
En este tipo de modelos, principalmente la curva de costo de combustible está dada por funciones de la potencia generada tipo polinomios [126, 128, 140]:

$$f(p) = a_0 + \sum_{j=1}^{M} a_j p^j. \tag{4}$$

Para una mayor facilidad a la hora de resolver el problema (3), las Empresas de Generación (GENCOs) consideran el costo del combustible usualmente como una función lineal de la potencia generada, o como una función cuadrática convexa, i.e., $M \in \{1, 2\}$ y $a_2 \geq 0$, en (4).

Funciones costo de combustible no suaves
Sin embargo, en la operación real de las unidades generadoras, varios aspectos pueden afectar la suavidad de las funciones de costo de combustible. En la actualidad y parcialmente motivados por el aumento de las capacidades computacionales, los investigadores han venido representando funciones de costo de combustible mediante modelos más complejos que toman en cuenta, por ejemplo, los efectos de las válvulas de admisión de vapor. Este fenómeno ocurre en turbinas de

vapor con múltiples válvulas, produciendo una curva ondulada, lo que motiva la adición de un término sinusoidal [7]:

$$f(p) = a_0 + \sum_{j=1}^{M} a_j p^j + |b \sin\left(c(p_{\min} - p)\right)|.$$

Según [7], este tipo de modelo aumenta la precisión de los resultados, pero evidentemente agrega más carga computacional, especialmente al perderse la convexidad. En la Figura 4 podemos observar la forma de la función de costo de combustible cuando se consideran las válvulas reguladoras de vapor.

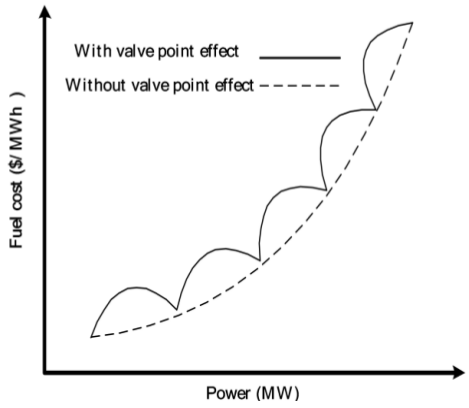


Figure 4: Curva de costo de combustible de una unidad generadora con válvulas de vapor [1].

Asimismo, las unidades generadoras pueden tener lo que se conoce como zonas de operación prohibidas. En este caso, se introducen discontinuidades en la formulación, ya que ahora existen zonas donde no se puede operar la unidad generadora, principalmente por aspectos técnicos, como posibles vibraciones del eje o por las propias características de las válvulas de vapor. En la Figura 5, se muestra una curva de costo de combustible para una unidad generadora con dos zonas de operación prohibidas.
Se pueden incorporar también otros elementos de costo a la función objetivo, dependiendo de las necesidades del tomador de decisiones (DM), por ejemplo:

- Costo de sobre-generación o infra-generación [80, 102]
- Costos de mantención y operación [91, 102]
- Costo de arrancada (en frío o caliente) [102]
- Costo de rampa [76]
- Costo de reserva de giro [76]
- Costo de transmisión [65]

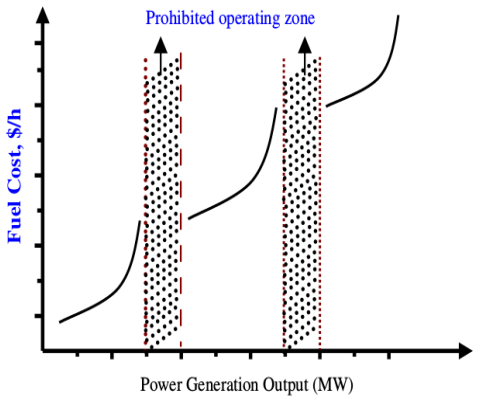


Figure 5: Curva de costo de combustible de una unidad generadora con zonas de operación prohibidas [94].

Restricciones del problema

Las redes de energía eléctrica son sistemas muy complejos. Las topologías de dichas redes pueden ser muy diferentes, pero en general existen varios agentes que deben coordinarse además de las GENCOs: las empresas de transmisión (TRANSCOs) y las empresas de distribución (DISTCOs) [154]. En la Figura 6 podemos observar una red de energía bien simple pero ilustrativa. La energía eléctrica producida por las GENCOs es transmitida por las TRANCOs, que poseen y operan las líneas de transmisión. Las DISTCOs por su parte poseen y operan las líneas de distribución, que es la etapa final en la entrega de energía eléctrica; estas empresas son las encargadas de llevar electricidad desde el sistema de transmisión a los consumidores individuales. Esto se realiza a través de subestaciones de distribución que están conectadas al sistema de transmisión y tienen como objetivo disminuir la tensión (voltaje) de transmisión a los valores medios y bajos de tensión requeridos por los consumidores finales. Esta complejidad se incorpora al problema ED a través de las restricciones consideradas, que representan los límites técnicos y operativos de todo el sistema [3].

1. Restricciones de generación

Balance de carga

Una particularidad de los sistemas de redes eléctricas es que la energía eléctrica no se puede almacenar, todavía, en grandes cantidades. Por ello, la generación de energía debe ser igual a la demanda, más las pérdidas de potencia en la línea de transmisión:

$$1^\top p - P_D - P_L = 0,$$

donde $p = (p_1, p_2, \ldots, p_N)^\top$ es el vector de potencias de cada unidad generadora; P_D es la demanda de energía y P_L las pérdidas en la transmisión. Las pérdidas de potencia se obtienen resolviendo el problema de flujo de carga, donde se tienen restricciones de igualdad para las potencias activas y reactivas en cada barra [103]:

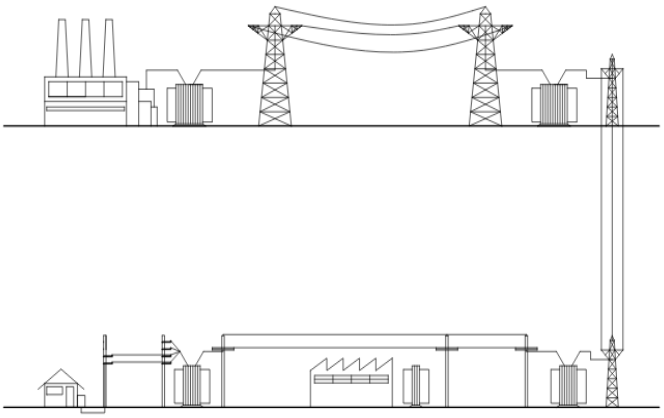


Figure 6: Red de potencia eléctrica.

$$P_{G_i} = P_{D_i} + V_i \sum_{j=1}^{NB} V_j [G_{ij} \cos(\delta_i - \delta_j) - B_{ij} \sin(\delta_i - \delta_j)],$$

$$Q_{G_i} = Q_{D_i} + V_i \sum_{j=1}^{NB} V_j [G_{ij} \sin(\delta_i - \delta_j) + B_{ij} \cos(\delta_i - \delta_j)],$$

donde NB es el número de barras; P_{G_i} y Q_{G_i} son la potencia real y reactiva generada en la barra i; P_{D_i} y Q_{D_i} son la carga de potencia real y reactiva en la barra i, respectivamente; G_{ij} y B_{ij} son la conductancia de transferencia y la susceptancia de transferencia entre la barra i y la j respectivamente; V_i y V_j son las magnitudes de voltaje en la barra i y j respectivamente; δ_i y δ_j son los ángulos de voltaje en la barra i y j respectivamente. Existen varios métodos para resolver el sistema no lineal resultante, entre los más populares se encuentra el método de Newton-Raphson. La solución de flujo de carga proporciona todas las magnitudes y ángulos de voltaje de la barra y se pueden usar para calcular las pérdidas de transmisión de la siguiente manera:

$$P_L = \sum_{k=1}^{NL} g_k [V_i^2 + V_j^2 - 2V_i V_j \cos(\delta_i - \delta_j)],$$

donde NL es el número de líneas de transmisión y g_k es la conductancia de la línea k conectando la barra i a la j. Otros autores calculan las pérdidas utilizando la fórmula de coeficientes de Kron [101]:

$$P_L = \sum_{i=1}^{N} \sum_{j=1}^{N} p_i B_{ij} p_j + \sum_{i=1}^{N} B_{0_i} p_i + B_{00},$$

donde p_i, p_j son las potencias de los generadores i y j respectivamente y B, B_0, B_{00} son coeficientes de pérdida.

Límites de generación
Cada generador debe producir energía dentro de límites inferiores y superiores, límites que se establecen de acuerdo con parámetros técnicos específicos:

$$p_{\min} \leq p \leq p_{\max}.$$

Zonas de operación prohibidas
Como se discutió anteriormente, en algunas unidades generadoras no siempre está disponible todo el rango operativo para la asignación de carga. Estas zonas de operación prohibida introducen discontinuidades en la función de costo que se modelan como [112]:

$$\begin{aligned} p_i^{\min} \leq \;& p_i \;\leq\; p_{i,1}^{L}, \\ p_{i,k-1}^{U} \leq \;& p_i \;\leq\; p_{i,k}^{L},\ k = 2, ..., n_i, \\ p_{i,n_i}^{U} \leq \;& p_i \;\leq\; p_i^{\max}, \end{aligned}$$

donde $p_i^{\min}$ y $p_i^{\max}$ son los límites de generación mínimos y máximos de la unidad i, respectivamente; n_i es el número de zonas de operación prohibidas; k indica la zona de operación prohibida de una unidad; $p_{i,k}^{L,U}$ son las fronteras inferior y superior de la zona prohibida k correspondiente a la unidad i.

2. Restricciones de seguridad

Flujo de línea
La restricción de flujo de línea se utiliza para evitar cargas de línea no deseadas debido a la distribución de energía. Por lo tanto, la carga de la línea de transmisión S_l está restringida por su límite superior como [113]:

$$S_{l_i} \leq S_{l_i}^{\max}, \quad i = 1, 2, ..., N_l,$$

donde N_l es el número de líneas del sistema.

Flujo de potencia aparente
A veces también es necesario limitar el flujo de potencia aparente [146]:

$$|LF_{f_i}| \leq LF_{f_i}^{\max}, \quad i = 1, 2, ..., N_T,$$

donde LF_{f_i} es el flujo de potencia aparente en la línea f_i; $LF_{f_i}^{\max}$ es el límite máximo para el flujo de energía aparente en la línea f_i y N_T es el número total de líneas de transmisión.

Restricción de reserva de giro del sistema
Para lograr una operación confiable y segura en los sistemas de generación de energía, es habitual considerar una reserva de (en) giro. Esta reserva de energía se guarda por si sucede alguna condición de operación no anticipada, como una interrupción inesperada de las unidades generadoras o un aumento repentino de la demanda. En [163] esta restricción se modela como:

$$\sum_{i=1}^{N} SP_{G_i} \geq SP_R, \qquad SP_{G_i} = \begin{cases} 0, & \forall i \in \Omega \\ P_{G_{\max_i}} - P_{G_i}, & \text{otro} \end{cases}$$

donde Ω es el conjunto de generadores con zonas de operación prohibidas, SP_{G_i} es la contribución de reserva en giro del generador i; SP_R es el requerimiento de reserva en giro del sistema; N es el

número total de unidades generadoras y P_{G_i},$P_{G_{\max_i}}$ son la potencia activa y su límite máximo para el generador i, respectivamente.

Despacho Económico y Ambiental

Por otro lado, el acelerado deterioro ambiental y sus marcados efectos ha motivado el desarrollo de sistemas de producción de energía más respetuosos con el medio ambiente. Según [118], desde los años setenta el estudio de la contribución de la industria energética al deterioro de su entorno se ha convertido en una importante tarea social.
La generación de energía mediante combustibles fósiles afecta significativamente el medio ambiente y la salud humana [68]. Durante la combustión se liberan emisiones de dióxido de azufre (SO_2), óxidos de nitrógeno (NO_x), material particulado (PM), monóxido de carbono (CO), compuestos orgánicos volátiles (COV), entre otros elementos. Estos son expulsados al aire a través de chimeneas, afectando zonas aledañas a las unidades generadoras. Los efectos ambientales de los contaminantes mencionados anteriormente, incluyen la reducción de la visibilidad, daños a los materiales, daños a la vegetación, deposición en forma de lluvia ácida, formación de ozono y contribución al efecto invernadero [100]. Por otro lado, en la Tabla 2, se presentan los principales contaminantes del aire y las posibles afecciones asociadas.

Table 2: Contaminantes y sus efectos [159].

Contaminante	Enfermedades asociadas
Material Particulado respirable	Enfermedades respiratorias, incluyendo bronquitis crónica y asma, enfermedades cardíacas
SO_2	Enfermedades cardíacas, efisémas pulmonares, cáncer, dolores de cabeza, irritación ocular
NO_2	Irritación pulmonar, infecciones virales, opresión del pecho
Material Particulado suspendido	Neumoconiosis, asma, cáncer
Ozono	Deterioro de la función pulmonar, dolor en el pecho, tos, irritación nasal y ocular
CO	Desmayos, muerte por asfixia

Por lo tanto, los problemas de las emisiones generadas por centrales eléctricas a base de combustibles fósiles, no pueden pasarse por alto. Según [95], muchas estrategias, como la instalación de filtros de aire y equipos de limpieza de contaminantes, el cambio a combustibles de bajas emisiones, la sustitución de equipos y unidades generadoras antiguas y la integración de fuentes de energía renovable, se han tenido en cuenta para minimizar las emisiones. Pero todos estas variantes necesitan cantidades considerables de capital y van asociadas a estrategias de reemplazo/modificación/actualización, por lo que se pueden considerar como soluciones a largo plazo. Por lo tanto, las estrategias de despacho de emisiones (donde las preocupaciones ambientales se adicionan a la formulación tradicional del ED) resultan soluciones atractivas, ya que son fáciles de implementar y requieren menos costos adicionales [50,143]. Usualmente para la representación de las emisiones se utilizan varios modelos. En la mayoría, el CO_2 se considera una función lineal de la potencia generada y el SO_x se considera proporcional al consumo de combustible de la unidad generadora, por lo tanto con una función de emisiones con la misma forma de (4). Por otro lado, las emisiones de NO_x son más difíciles de evaluar porque provienen de dos fuentes diferentes: del nitrógeno en el aire y en el combustible; además su producción está relacionada con varios factores, incluida la temperatura de la caldera y el contenido de aire. Por lo tanto, la curva de salida de NO_x se puede caracterizar por una ecuación que consta de funciones cuadráticas o lineales, más términos exponenciales adicionales [103, 133]. Por tanto, atendiendo al tipo de contaminante a evaluar, debemos considerar la emisión total como:

$$E(p) \doteq \sum_{i=1}^{N}\sum_{j=1}^{M} \alpha_{0i} + \alpha_{ji} p_i^j, \text{ ó}$$

$$E(p) \doteq \sum_{i=1}^{N}\sum_{j=1}^{M} \alpha_{0i} + \alpha_{ji} p_i^j + \xi_i e^{\lambda_i p_i},$$

donde α_{ji}, ξ_i y λ_i son los coeficientes de emisión de la unidad generadora i.

En el Capítulo 1 se analiza el problema del Despacho Económico Ambiental:

$$\begin{aligned} &\min \ (F(p), E(p)) \\ &st \quad 1^\top p = P_D, \\ &p_{\min} \leq p \leq p_{\max}, \end{aligned} \tag{5}$$

donde, $F(p) \doteq \sum_{i=1}^{N}\sum_{j=1}^{M} \hat{a}_{0i} + \hat{a}_{ji} p_i^j$ es el costo total de combustible; $E(p) \doteq \sum_{i=1}^{N}\sum_{j=1}^{M} \hat{\alpha}_{0i} + \hat{\alpha}_{ji} p_i^j$ son las emisiones totales y además se considera un sistema sin pérdidas en la transmisión. Se propone un método de solución basado en teoría de dualidad y se comprueban los resultados propuestos en varios sistemas. Las soluciones que se obtienen dominan las encontradas por otros autores utilizando metaheurísticas. Este capítulo se basa fundamentalmente en el artículo [30]:

- Adrian CARRILLO-GALVEZ, Fabián FLORES-BAZÁN, Enrique LÓPEZ: A duality theory approach to the environmental/economic dispatch problem. Electric Power Systems Research, vol 184, Art. Num. 106285, (2020).

A su vez, dicho artículo se compone de los siguientes trabajos previos, los cuales fueron presentados y publicados en [31, 32]:

- Adrian CARRILLO-GALVEZ, Fabián FLORES-BAZÁN, Enrique LÓPEZ: An Analytical Approach to the Environmental/Economic Dispatch Problem. 2019 IEEE CHILEAN Conference on Electrical, Electronics Engineering, Information and Communication Technologies (CHILECON), (2019), pp. 1-5, doi: 10.1109/CHILECON47746.2019.8987495.
- Adrian CARRILLO-GALVEZ, Fabián FLORES-BAZÁN, Enrique LÓPEZ: On the solution of the Environmental/Economic Dispatch problem using Lagrangian duality. 2020 IEEE International Conference on Industrial Technology (ICIT), (2020), pp. 619-623, doi: 10.1109/ICIT45562.2020.9067261.

Sin embargo, cuando se analiza la literatura existente, se puede evidenciar que existe un problema estrechamente relacionado al despacho de emisiones y que es sistemáticamente ignorado. Debe tenerse en cuenta que, generalmente al resolver (5), todos los coeficientes de las funciones de emisión y costo de combustible se consideran valores fijos (exactamente conocidos). Pero este enfoque determinístico puede conducir a errores significativos en el programa de generación óptimo propuesto. Consideremos que, en la operación real de las redes de energía eléctrica, las empresas generadoras deben establecer el costo de combustible y las emisiones de manera experimental. Esto se puede

hacer de diversas maneras: a través de pruebas de rendimiento, la determinación a partir de registros operativos o el uso de los datos de la garantía del fabricante, ajustados a las condiciones reales de funcionamiento [78].
Pero, el alto costo de las pruebas de desempeño y la posible representación incorrecta de la forma de las curvas de entrada-salida proporcionada por los fabricantes, hace que la determinación a partir de los registros operativos sea la alternativa más empleada; más aún si tenemos en cuenta que, en la actualidad, los datos operativos están fácilmente disponibles en muchas empresas de generación. Esta estrategia claramente introduce incertidumbres en los modelos derivados experimentalmente, incertidumbres que pueden llevarnos a obtener soluciones dominadas (si se considera el problema EED en su formulación determinísitica (5)); o a soluciones ambientales inviables (si se considera la formulación determinísitica para el ECED). En el Capítulo 2 analizamos el efecto que tiene las incertidumbres en las estimaciones de los coeficientes. Comparamos los programas de generación óptimos obtenidos al resolver los problemas determinísticos o utilizando las formulaciones propuestas. Los enfoques propuestos se basan en intervalos estadísticos, particularmente en los respectivos intervalos de predicción involucrados cuando se utiliza la regresión lineal múltiple para ajustar los modelos. De hecho, dado que se tiene la suma de variables respuesta, se debe obtener un intervalo de predicción para esta suma. La formulación propuesta también fue comprobada en un sistema real, consistente en 4 unidades generadoras del sistema eléctrico de Chile. Los resultados obtenidos muestran las afectaciones que pudieran tenerse cuando no se consideran las incertidumbres en las estimaciones de los coeficientes de costo y emisiones. Estas afectaciones pueden ser graves, sobre todo durante el invierno, cuando la mala ventilación en diversas zonas del país disminuye la calidad del aire. Los resultados de este capítulo se basan principalmente en:

- Adrian CARRILLO-GALVEZ, Fabián FLORES-BAZÁN, Enrique LÓPEZ: Effect of Models Uncertainties on the Emission Constrained Economic Dispatch. A Prediction Interval-Based Approach. **Sometido para publicación.**

Los resultados contenidos en la segunda parte de esta tesis resultan más generales que los obtenidos en la primera parte. Aunque fueron desarrollados y motivados principalmente por un problema específico, como podemos ver en el Capítulo 3, los resultados obtenidos fueron (son) de amplia aplicación y trascienden la mera gestión de las emisiones y costos de redes eléctricas. Consideremos que, como se señaló en [116], los sistemas y mercados energéticos modernos están envueltos en diversas fuentes de incertidumbre, debido a la revolución de las energías renovables (ER) en los últimos años. El valor de las inversiones en tecnologías renovables ha aumentado rápidamente durante la última década, como resultado de las presiones políticas para reducir las emisiones de dióxido de carbono y los incentivos políticos para aumentar la participación de las energías renovables en la matriz energética. A medida que aumenta el número de inversiones en energía renovable, también lo hace la necesidad de medir los riesgos asociados a la planificación, construcción y operación de estas tecnologías. Los desarrolladores del sector de la energía, los inversores y los responsables políticos se enfrentan a un futuro que implica implícitamente riesgos e incertidumbres tecnológicas, financieras y políticas. Aunque, las tecnologías de ER tienen potencialmente un perfil de riesgo más bajo que las fuentes de energía convencionales, porque están desconectadas de los precios de los combustibles fósiles; aún conllevan una considerable exposición al riesgo tecnológico, financiero y regulatorio, dependiendo de la tecnología, el país y el régimen regulatorio imperante. La fluctuación de los componentes del costo de las unidades de generación de energía, los precios volátiles del petróleo, el precio de la electricidad y el costo del carbono en el contexto de la estrategia global

de mitigación del cambio climático, son ejemplos de componentes de incertidumbre que encuentran los desarrolladores, inversionistas y legisladores en el sector energético [79]. Por otro lado, como se señala en [19], un gran inconveniente de las energías renovables en un mercado eléctrico desregulado es su alto precio, lo que reduce su competitividad. A pesar de lo anterior, la energía renovable sigue siendo ventajosa y sobre todo una estrategia muy adecuada para disminuir las emisiones contaminantes. En este contexto los análisis y resultados provenientes de la Teoría de Portafolios, proporcionan una herramienta útil para la gestión eficiente de los riesgos e incertidumbres asociadas a la planificación y operación de sistemas de energía sostenibles, ver por ejemplo [79, 116] ó [117]. En general, la teoría clásica de portafolios (considerando medias y varianzas) [97, 98] se basa en un problema de optimización, que puede ser visto como una extensión del problema estándar cuadrático (StQOP):

$$\mu_q \doteq \min\Big\{\frac{1}{2}x^\top Ax:\ e^\top x = 1,\ x \in C\Big\},$$

donde A es una matriz real y simétrica de orden n; $C \subseteq \mathbb{R}^n$ es un cono convexo, cerrado y puntiagudo que tiene un interior no vacío y $e \in \text{int } C^*$, siendo C^* el cono polar no negativo de C. En el Capítulo 3 no sólo analizamos el Problema del Portafolio, sino una amplia clase de problemas de minimización donde las funciones objetivo y la restricción son positivamente homogéneas de grados posiblemente diferentes y, además, las restricciones geométricas vienen dadas por un cono convexo (no necesariamente poliédrico); es decir, discutimos y analizamos el problema de optimización homogénea no convexo:

$$\min\{f(x):\ g(x) = 1,\ x \in C\},$$

donde $C \subseteq \mathbb{R}^n$ es un cono cerrado (no necesariamente convexo) y f, g son funciones positivamente homogéneas en C, con diferentes grados de homogeneidad y además $g(x) > 0$ para toda $x \in C$, $x \neq 0$. Varias caracterizaciones son brindadas acerca de la validez de la propiedad de dualidad fuerte respecto al problema dual Lagrangiano asociado. Una de las caracterizaciones está relacionada con la convexidad de $(g, f)(C) + \mathbb{R}_+(0, 1)$, revelando cierta convexidad escondida y un S-lema. Cuando ambas funciones tienen el mismo grado de homogeneidad, se establece un formulación copositiva del problema original. También son derivadas condiciones de optimalidad de orden cero y condiciones KKT (locales o globales) que originan la noción de L-valores propios con aplicaciones además al análisis de tensores simétricos. Se estudió particularmente el problema cuando el cono C está dado por dos formas cuadráticas. Estos resultados fueron aplicados también a una clase de problemas de optimización fraccionaria cuadrática con dos restricciones cuadráticas, obteniéndose nuevas condiciones de optimalidad de segundo orden, tanto necesarias como suficientes. Los resultados obtenidos en este capítulo se basan principalmente en [29]:

- Adrian CARRILLO-GALVEZ, Fabián FLORES-BAZÁN: Non convex homogeneous optimization: a general framework and applications to quadratic fractional programming and tensor eigenvalues analysis. **Sometido para publicación**. (Preprint disponible en: https://www.ing-mat.udec.cl).

Part I. Combined Economic Emission Dispatch

Chapter 1

A duality theory approach to the EED

1.1 Introduction

Increasing concerns about global warming and environmental deterioration, has drawn more attention on daily optimal operation of electric power systems. Several strategies exist to mitigate both, production and emission of pollutants during the generation of electrical energy, but Environmental Economic Power Dispatch or Environmental/Economic Dispatch (EED) has proved to yield attractive solutions [95]; specially for electrical power networks, as the existing in developing countries, with low penetration of renewable energy sources. Thus, the EED becomes an extension of the traditional ED problem, when environmental concerns are added in the usual mathematical formulation.

Numerous investigations have dealing with the EED problem (and similar problems derived from it), but have been principally addressed to the solution methods and numerical results (comprehensive reviews of the principal results and used techniques can be seen in [37,96,123]). These papers show a clearly predominance of heuristic methods that are capable of achieve satisfactory but suboptimal solutions. However, and this partially motivated our investigation, there are fewest papers focused on theoretical analysis and giving some insights of the analytical treatment of the problem. The principal reasons for this lack of analytical research is that the employed methods, in general, need multiple runs to find the Pareto front, so computational time becomes longer; and some convexity assumptions are needed. In this chapter we develop a duality theory approach to the EED problem that handle these complexities. Due to the convexity of the Pareto front for the EED problem, we obtained the whole Pareto front using the Weighted Sum Method (WSM). The structure of the scalar problem resulting from the application of the WSM ensure, in a relatively easy way, strong duality property to the primal problem. Therefore, we recover the solution to the EED by solving the dual problem, which becomes a quadratic programming problem simplest than the original and therefore with lower compilation times.

The remainder of the chapter is organized as follows: Section 1.2 is intended to introduce some mathematical aspects about optimization theory, specially multiobjective optimization and duality theory; in Section 1.3 the EED problem is analyzed and the duality theory approach is developed; in Section 1.4 the solution method is tested on three study cases and the results are compared with previous papers; in Section 1.5 we introduce a heuristic algorithm providing better distributed non-dominated solutions; a comparison with some analytical solutions available appears in Section 1.6 and Section 1.7 concludes the chapter.

1.2 Notation and basic optimization theory

For a better and complete understanding of the problem and future ideas, some previous theoretical aspects are needed. Let us begin with some notation and definitions that will be used through this chapter:

- The real line is denoted by $\mathbb{R}$ and the p-dimensional Euclidean space is denoted by $\mathbb{R}^p$.
- The set of all non-negative vectors of $\mathbb{R}^p$ is denoted by $\mathbb{R}^p_+$, i.e., $\mathbb{R}^p_+ \doteq \{x \in \mathbb{R}^p \ : x_i \geq 0, \ \forall i = 1, 2, \ldots, p \ \}$.
- When strict positivity holds we define $\mathbb{R}^p_{++} \doteq \{x \in \mathbb{R}^p \ : x_i > 0 \ , \ \forall i = 1, 2, \ldots, p\}$.
- $\langle \cdot, \cdot \rangle$ represent the usual scalar product.
- $e_i \in \mathbb{R}^p$ denotes the ith vector of the canonical base of $\mathbb{R}^p$.
- I_p is the identity matrix of dimension $p \times p$.
- $A \succeq 0$ means the matrix A is positive semidefinite, i.e., $x^\top Ax \geq 0$ for all x.

1.2.1 Multiobjective optimization

A multiobjective optimization problem (MOP) in that the decision-maker wants to optimize k objectives, can be stated, in a general form, by:

$$\min_{x \in \mathcal{X}} \ (f_1(x), \ldots, f_k(x)), \tag{1.1}$$

where x is a vector from $\mathbb{R}^p$ and $\mathcal{X}$ is referred as the *feasible set* or the set of alternatives of the decision problem. This feasible set is determined by the set of constraints, from now on $g_i : \mathbb{R}^p \longrightarrow \mathbb{R}$ and it can be of equality or inequality type. Additionally the set $\mathcal{Y} = f(\mathcal{X}) \doteq \{y \in \mathbb{R}^k : y_i = f_i(x) \ , \ \ i = 1, \ldots, k, \text{ for some } x \in \mathcal{X}\}$ is referred to as the *image of the feasible set* or the feasible space in the criterion space via the objective functions.
The different existing notions of solution for a MOP are based on the observation that, when we have multiple objectives to optimize, generally, the objective functions are non-commensurable and often competing and conflicting. This gives a set of optimal solutions instead of only one; and the type of the solutions is related to the form in that they are compared, since not total order in $\mathbb{R}^k$ exist. We will use the following notation [46]:

Notation 1.1. *For any $y^1, y^2 \in \mathbb{R}^k$:*

- $y^1 \leq y^2 \Leftrightarrow y^2_i \in [y^1_i, \infty), \text{for all i} = 1, \ldots, \text{k}, \ \text{y}^1 \neq \text{y}^2$.
- $y^1 \leqq y^2 \Leftrightarrow y^2_i \in [y^1_i, \infty), \text{for all i} = 1, \ldots, \text{k}$.

Let us see the next definition.

Definition 1.1. *[46] A feasible solution $\bar{x}$ is called efficient or Pareto optimal if there is no other $x \in \mathcal{X}$ such that $f(x) \leq f(\bar{x})$. If $\bar{x}$ is efficient, $f(\bar{x})$ is called a non-dominated point. If $x^1, x^2 \in \mathcal{X}$ and $f(x^1) \leq f(x^2)$, we say x^1 dominates x^2 and $f(x^1)$ dominates $f(x^2)$. The set of all efficient solutions is denoted by $\mathcal{X}_E$ and called the efficient set. The set of all non-dominated points $\bar{y} = f(\bar{x}) \in \mathcal{Y}$, where $\bar{x} \in \mathcal{X}_E$ is denoted $\mathcal{Y}_N$ and called the non-dominated set.*

There are other definitions of efficiency frequently used, and we shall refer often to one of which is best suited in a particular context. Equivalently $\bar{x} \in \mathcal{X}$ is *efficient* if:

- there is no $x \in \mathcal{X}$ such that $f_k(x) \leq f_k(\bar{x})$ for $k = 1, \ldots, p$, and $f_i(x) < f_i(\bar{x})$ for some $i \in \{1, \ldots, p\}$,
- there is no $x \in \mathcal{X}$ such that $f(x) - f(\bar{x}) \in -\mathbb{R}^p_+ \setminus \{0\}$,
- $f(\mathcal{X}) \cap (f(\bar{x}) - \mathbb{R}^p_+) = \{f(\bar{x})\}$.

Another notion is *weakly* efficient solution.

Definition 1.2. *A feasible solution $\bar{x}$ is called* weakly efficient *or* weakly Pareto optimal *if there is no other $x \in \mathcal{X}$ such that $f(x) < f(\bar{x})$, i.e. $f_k(x) < f_k(\bar{x})$ for all $k = 1, \ldots, p$. The point $\bar{y} = f(\bar{x})$ is then called* weakly non-dominated. *The weakly efficient and non-dominated sets are denoted $\mathcal{X}_{wE}$ and $\mathcal{Y}_{wE}$ respectively.*

From the definitions is obvious that:

$$\mathcal{X}_E \subset \mathcal{X}_{wE}.$$

Among the existing solution techniques to MOPs, perhaps the most known are the ϵ-Constrained Method and the already mentioned Weighted Sum Method.

ϵ-Constrained Method

In this method only one of the original objectives is minimized, while the others are transformed to constraints , i.e., problem (1.1) is substituted by the ϵ-constraint problem [46]:

$$\begin{array}{cc} (\epsilon) & \min_{x \in \mathcal{X}} \ f_j(x) \\ & st \quad f_k(x) \leq \epsilon_k, \ k = 1, \ldots, p, \quad k \neq j, \end{array} \tag{1.2}$$

where $\epsilon \in \mathbb{R}^p$. This method is quite popular between practitioners, since it is straightforward to employ one-objective optimization technique now. From the properties of problem (1.2) and its relationship with the efficient solutions of (1.1), we point out several theoretical results.

Proposition 1.1. *[46] Let $\bar{x}$ be an optimal solution of 1.2 for some j. Then $\bar{x} \in \mathcal{X}_{wE}$.*

Proof. By the opposite, assume $\bar{x} \notin \mathcal{X}_{wE}$. Then there is an $x \in \mathcal{X}$ such that $f_k(x) < f_k(\bar{x})$ for all $k = 1, \ldots, p$. In particular, $f_j(x) < f_j(\bar{x})$. Since $f_k(x) < f_k(\bar{x}) \leq \epsilon_k$ for $k \neq j$, the solution x is feasible for (1.2). This is a contradiction to $\bar{x}$ being an optimal solution of (1.2). □

In order to strengthen the above proposition to obtain efficiency we require the optimal solution of (1.2) to be unique.

Proposition 1.2. *[46] Let $\bar{x}$ be a unique optimal solution of 1.2 for some j. Then $\bar{x} \in \mathcal{X}_E$.*

Proof. By the opposite, assume there is some $x \in \mathcal{X}$ with $f_k(x) \leq f_k(\bar{x}) \leq \epsilon_k$ for all $k \neq j$. If in addition $f_j(x) \leq f_j(\bar{x})$ we must have $f_j(x) = f_j(\bar{x})$ because $\bar{x}$ is an optimal solution of (1.2). So x is an optimal solution of (1.2). Thus, uniqueness of the optimal solution implies $x = \bar{x}$ and $\bar{x} \in X_E$. □

In general with appropriate choices of ϵ all efficient solutions can be found; however a previously problem must be solve in order to provide minimum and maximum value of each individual objective function.

Weighted Sum Method

On another hand the WSM is based on the idea of solve (1.1), by solving instead a single objective problem of the type:

$$(WS) \qquad \min_{x \in \mathcal{X}} \sum_{k=1}^{p} w_k f_k(x). \tag{1.3}$$

Several important results exist for a vector $\bar{x}$ solution of Problem (1.3) and its relationship with (1.1).

The next propositions show us that optimal solutions of the weighted sum problem with positive (nonnegative) weights are always (weakly) efficient and that under convexity assumptions all (weakly) efficient solutions are optimal solutions of scalarized problem with positive (nonnegative) weights. The proofs are elementary and can be seen in [46].

Proposition 1.3. *[46] Suppose that $\bar{x}$ is an optimal solution of the weighted sum optimization problem:*

$$\min_{x \in \mathcal{X}} \sum_{k=1}^{p} w_k f_k(x),$$

with $w \in \mathbb{R}^p_+$. Then the following statements hold:

- *If $w \in \mathbb{R}^p_+$ then $\bar{x} \in \mathcal{X}_{wE}$,*
- *If $w \in \mathbb{R}^p_{++}$ then $\bar{x} \in \mathcal{X}_E$,*

To have results in the other direction, convexity assumptions are needed.

Proposition 1.4. *[46] Let $\mathcal{X}$ be a convex set, and let f_k be convex functions, $k = 1, \ldots, p$. If $\bar{x} \in \mathcal{X}_{wE}$ there is some $w \in \mathbb{R}^p_+$ such that $\bar{x}$ is an optimal solution of* (1.3)*.*

Particularly, in our investigation the WSM was selected over the ϵ-Constrained Method since, as we will see later in this chapter:

- For the EED problem analyzed in here we have the convexity of the set $\mathcal{Y}$. This follows from the technical characteristics of most thermal generating units.
- In the ϵ-Constrained Method one of the objective function is treated as a restriction, introducing an extra Lagrange multiplier when the scalar optimization problem is solve. However, for different fuel cost and emission functions, this method overcame the main disadvantage of the WSM when convexity does not hold.

1.2.2 Duality theory

Consider a nonlinear programming problem (1.4), which will be referred as the primal problem:

$$\begin{aligned} & \mu \doteq \min \ f(x) \\ \text{st} \quad & g_i(x) \leq 0 \quad \text{for } i = 1, \ldots, m, \\ & h_i(x) = 0 \quad \text{for } i = 1, \ldots, l, \\ & x \in \mathcal{X}. \end{aligned} \tag{1.4}$$

There exist several closely related problems associated to this primal one in literature, known as dual problems. Perhaps the most usual dual formulation is the Lagrangian:

$$\begin{aligned} & \nu \doteq \max \ \theta(\lambda_1, \lambda_2), \\ \text{st} \quad & \lambda_1 \in \mathbb{R}^l, \ \lambda_2 \in \mathbb{R}^m_+ \end{aligned} \tag{1.5}$$

where:

$$\theta(\lambda_1, \lambda_2) \doteq \inf_{x \in \mathcal{X}} \{f(x) + \lambda_1^\top h(x) + \lambda_2^\top g(x)\}.$$

Clearly, for a nonlinear programming problem, several Lagrangian dual problems can be obtained, depending on which constraints are handled as $g(x) \leq 0$ and $h(x) = 0$ and which constraints are treated by the set $\mathcal{X}$. This choice can affect both, the optimal value of (1.5) (as in non-convex situations) and the effort expended in evaluating and updating the dual function θ, during the course of solving the dual problem. Hence, an appropriate selection of the set $\mathcal{X}$ must be made, depending on the structure of the problem and the purpose for solving (1.5) [16]. Now we present some basic results and definitions about the relationship between primal and dual problems. Further (broader) results are in Chapter 3.

Theorem 1.1 (Weak Duality Theorem [16])**.** *Let $\bar{x}$ be a feasible solution to Problem* (1.4)*; that is, $\bar{x} \in \mathcal{X}$, $g(\bar{x}) \leq 0$ and $h(\bar{x}) = 0$. Also let (λ_1, λ_2) be a feasible solution to problem* (1.5)*, this is, $\lambda_2 \geq 0$. Then $f(\bar{x}) \geq \theta(\lambda_1, \lambda_2)$.*

The proof is simple and can be viewed for example in [16]. From this result it can be proved that the objective value of any feasible solution to the dual problem yields a lower bound on the objective value of any feasible solution to the primal problem. In other words, the optimal value of the primal problem is greater than or equal to the optimal value of the dual problem. If strict inequality holds true ($\nu < \mu$), a *duality gap* is said to exist. It is said that (1.4) has the *strong-duality property* if the duality gap is zero ($\nu = \mu$) and problem (1.5) admits any solution [16]. Therefore, if the structure of the problem ensures strong duality property, Lagrangian duality becomes an attractive method for non-linear programming. Note that $\theta(\lambda_1, \lambda_2)$ is a pointwise minimum of affine functions, then a concave function. This way maximizing $\theta(\lambda_1, \lambda_2)$ over $\lambda_2 \in \mathbb{R}^m_+$ is a convex optimization problem, with less constraints than the primal problem.
There exist several theoretical results that give us conditions ensuring strong duality, but we remark (for the applicability to EED problem and its theoretical simplicity) the following:

Theorem 1.2 (Strong Duality Theorem [16]). *Let $\mathcal{X}$ be a nonempty convex set in $\mathbb{R}^n$, let $f : \mathbb{R}^n \longrightarrow \mathbb{R}$ and $g : \mathbb{R}^n \longrightarrow \mathbb{R}^m$ be convex, and let $h : \mathbb{R}^n \longrightarrow \mathbb{R}^l$ be affine; that is, h is of the form $h(x) = Ax - b$. Suppose that the following constraint qualification holds true. There exists an $\bar{x} \in \mathcal{X}$ such that $g(\bar{x}) < 0, h(\bar{x}) = 0$ and $0 \in \text{int } h(\mathcal{X})$, where $h(\mathcal{X}) \doteq \{h(x) : x \in \mathcal{X}\}$, then:*

$$\inf\{f(x) : x \in \mathcal{X},\ g(x) \leq 0,\ h(x) = 0\} = \sup_{\substack{\lambda_1 \in \mathbb{R}^l \\ \lambda_2 \in \mathbb{R}^m_+}} \{\theta(\lambda_1, \lambda_2)\}.$$

This theorem shows that under suitable convexity assumptions and under a constraint qualification, the optimal objective function values of the primal and dual problems are equal, so no duality gap exists. As we will see in the next section, particularly for the EED problem becomes relatively easy to ensure strong duality property.

1.3 EED mathematical analysis

In this section we analyze the EED problem and the properties that made a duality approach feasible. We start formulating the problem as a MOP.

1.3.1 Mathematical formulation

As it was previously stated, the Environmental/Economic Dispatch problem has as objectives to minimize the emissions of pollutants and the total fuel cost of meeting the energy requirements of the system. In this chapter we are going to deal with the following objective functions and constraints:

Objective functions

1. Objective function $F(p)$ aims to minimize the total fuel cost for the entire system:

$$\min F(p) = \min \sum_{i=1}^{N} f_i(p_i),$$

where:

- $p \in \mathbb{R}^N$ is the decision vector that denotes the power supply by each generating unit and N is the number of generating units of the system.
- $f_i(p_i)$ is the fuel cost function, in \$/h, for the ith unit, which is modeled by a quadratic polynomial function:

$$f_i(p_i) = a_i p_i^2 + b_i p_i + c_i,$$

with a_i, b_i, c_i, the fuel cost coefficients of unit i.

2. Objective function $E(p)$ describes the total quantity of emissions released to the environment by the operation of N units:

$$\min E(p) = \min \sum_{i=1}^{N} e_i(p_i),$$

where:

- $e_i(p_i)$ is the emission function, in kg/h, for the ith unit, which we particularly modeled, as well, by a quadratic polynomial function:

$$e_i(p_i) = \alpha_i p_i^2 + \beta_i p_i + \gamma_i,$$

with $\alpha_i, \beta_i, \gamma_i$, the emission coefficients of unit i.

Constraints

1. The real power output limits are related with the technical capacities of each generating unit:

$$p_{\min_i} \leq p_i \leq p_{\max_i}, \qquad i = 1, 2, \ldots, N,$$

where:

- $p_{\min_i}, p_{\max_i}$ are the minimum and maximum output limit of the ith generating unit in MW.

2. The real power balance characterize the fact that, all the demanded load must be supplied, instantly, by the electric power network. In a system this is modeled as:

$$\sum_{i=1}^{N} p_i = P_D + P_L,$$

where:

- P_D is the total load demand of the power system in MW.
- P_L is the system losses.

We will not considerate power losses for the development of the proposed approach. Therefore, in matrix notation, the problem to solve is formulated as a multiobjective constrained problem:

$$\begin{aligned} & \min \ (F(p), E(p)) \\ \text{st} \quad & 1^\top p = P_D, \\ & 0 \leq p_{\min} \leq p \leq p_{\max}, \end{aligned}$$

where:

1. $F(p)$ is the total fuel cost function:

$$F(p) = p^\top A p + b^\top p + c,$$

where:

- $A \doteq \text{diag}(\text{a}_1, \ldots, \text{a}_\text{N}),\ \text{a}_\text{i} > 0,\ \text{i} = 1, 2, \ldots, \text{N},$
- $b \doteq (b_1, b_2, \ldots, b_N)^\top$.
- $c \doteq \sum_{i=1}^{N} c_i$.

2. $E(p)$ is the function of the total emission of pollutants:

$$E(p) = p^\top Q p + r^\top p + s,$$

where:

- $Q \doteq \text{diag}(\text{q}_1, \ldots, \text{q}_\text{N}),\ \text{q}_\text{i} > 0,\ \text{i} = 1, 2, \ldots, \text{N},$
- $r \doteq (\beta_1, \beta_2, \ldots, \beta_N)^\top$.
- $s \doteq \sum_{i=1}^{N} \gamma_i$.

3. $p_{\min}, p_{\max} \in \mathbb{R}^N$ are vectors whose ith component denotes, respectively, the minimum and maximum output of power of the generator i and 1 is the all ones vector of N elements.

1.3.2 Mathematical structure analysis

Clearly, to minimize $F(p) = p^\top A p + b^\top p + c$ is equivalent to minimize $\bar{F}(p) = \frac{1}{2}(p^\top A p + b^\top p)$, similarly from $E(p)$ we have the equivalent objective function $\bar{E}(p) = \frac{1}{2}(p^\top Q p + r^\top p)$. Then, for simplicity, we are going to be interested in to solve the MOP:

$$\begin{aligned} (\bar{P}) \quad \min \quad & \tfrac{1}{2}\left(p^\top A p + b^\top p, p^\top Q p + r^\top p\right) \\ \text{st} \quad & 1^\top p = P_D, \\ & 0 \leq p_{\min} \leq p \leq p_{\max}, \end{aligned}$$

with A, Q both symmetric and positive semidefinite. This result is evident once that we realize both matrices are diagonal and with positive entries. Hence, both objective functions are convex, also continuous and differentiable, as well the constraint functions. Therefore (see Theorem 4.1 in [46]) we can compute the efficient solutions by varying the weight vector $w = (w_1, w_2)^\top \in \mathbb{R}^2_+$ and solving the scalarized problem:

$$\begin{aligned} \min \ & \frac{1}{2}[p^\top (w_1 A + w_2 Q) p + (w_1 b + w_2 r)^\top p] \\ \text{st} \quad & 1^\top p = P_D, \\ & 0 \leq p_{\min} \leq p \leq p_{\max}. \end{aligned} \tag{1.6}$$

Now we develop an approach based on duality theory to solve (1.6).

1.3.3 Duality theory approach

In this subsection the scalar problem (1.6) is solved by taking advantage of the structure of the EED problem, particularly the previous results that guarantee strong duality.
Let us rewrite the scalar problem as:

$$\mu \doteq \min \ \frac{1}{2}[p^\top (w_1 A + w_2 Q)p + (w_1 b + w_2 r)^\top p] \qquad (1.7)$$
$$\text{st} \quad 1^\top p = P_D,$$
$$Bp \leq c,$$

where:

$$B = \begin{pmatrix} -\mathrm{I}_N \\ \mathrm{I}_N \end{pmatrix} \ , \qquad c = \begin{pmatrix} -p_{\min} \\ p_{\max} \end{pmatrix}.$$

Let us define the Lagrangian function:

$$L(p, \lambda_1, \lambda_2) \doteq \frac{1}{2} p^\top (w_1 A + w_2 Q)p + \frac{1}{2}(w_1 b + w_2 r)^\top p + \ \lambda_1 (1^\top p - P_D) + \lambda_2^\top (Bp - c),$$

with $\lambda_1 \in \mathbb{R}, \lambda_2 \in \mathbb{R}_+^{2N}$; then we associate to the primal problem (1.7), the (Lagrangian) Dual Problem:

$$\max_{\substack{\lambda_1 \in \mathbb{R} \\ \lambda_2 \in \mathbb{R}_+^{2N}}} \ \inf_{p \in \mathbb{R}^N} L(p, \lambda_1, \lambda_2). \qquad (1.8)$$

If we define the dual function:

$$\theta(\lambda_1, \lambda_2) \doteq \inf_{p \in \mathbb{R}^N} \Big\{ \frac{1}{2} p^\top (w_1 A + w_2 Q)p + \frac{1}{2}(w_1 b + w_2 r)^\top p + \lambda_1 (1^\top p - P_D) + \lambda_2^\top (Bp - c) \Big\},$$

then we need to solve the Dual Problem:

$$\nu \doteq \max_{\substack{\lambda_1 \in \mathbb{R} \\ \lambda_2 \in \mathbb{R}_+^{2N}}} \ \theta(\lambda_1, \lambda_2). \qquad (1.9)$$

Clearly we can define several dual problems, depending on which constraints are included in the Lagrangian function or in the feasible set where $\theta(\lambda_1, \lambda_2)$ is searched. By choosing this specific dual formulation (1.8) we can:

1. To ensure, in a relatively easy way, strong duality. Therefore we can find μ by solving (1.9), which is easiest to solve since one less constraint must be considered.

Note that, in this formulation, we have:

- $f(p) \doteq \frac{1}{2}[p^\top (w_1 A + w_2 Q)p + (w_1 b + w_2 r)^\top p]$ is convex,
- $g(p) \doteq (g^1(p), g^2(p))^\top$ is convex; where $\forall i = 1, \ldots, N$, we have:
  - o $g_i^1(p) = -\langle p, e_i \rangle + \langle p_{\min}, e_i \rangle$,
  - o $g_i^2(p) = \langle p, e_i \rangle - \langle p_{\max}, e_i \rangle$,
- $h(p) \doteq 1^\top p - P_D$ is affine,

- $\mathcal{X} = \mathbb{R}^N$ is convex.

Therefore, conditions of Theorem 1.2 hold, even the constraint qualification. Note that automatically we have $0 \in \text{int } h(\mathcal{X})$. To see this we use the reasoning in [16]. Suppose that $h(p) = 1^\top p - P_D$, then clearly any $y \in \mathbb{R}$ can be represented as $y = 1^\top p - P_D$, with $p = (y + P_D, 0, \ldots, 0)$. This way $h(\mathcal{X}) = \mathbb{R}$ and particularly $0 \in \text{int } h(\mathcal{X})$.

2. To obtain $\theta(\lambda_1, \lambda_2)$ by solving a simple unconstrained Quadratic Programming problem in $\mathbb{R}^N$. Therefore we can obtain an analytic solution, by finding the point $\bar{p}$ in $\mathbb{R}^N$ such that $\frac{\partial L}{\partial p}\Big|_{p=\bar{p}} = 0$. So we need to solve the equation:

$$\frac{\partial L}{\partial p} = (w_1 A + w_2 Q)p + \frac{1}{2} w_1 b + \frac{1}{2} w_2 r + \lambda_1 1 + B^\top \lambda_2 = 0.$$

Hence,

$$\bar{p} = -(w_1 A + w_2 Q)^{-1}\Big(\frac{1}{2} w_1 b + \frac{1}{2} w_2 r + \lambda_1 1 + B^\top \lambda_2\Big). \tag{1.10}$$

If we substitute this result in (1.9), then:

$$\begin{aligned} \Theta(\lambda_1, \lambda_2) = \; & - \quad \lambda_1^2 1^\top \Sigma^{-1} 1 - \lambda_2^\top B \Sigma^{-1} B^\top \lambda_2 - 2\lambda_1 1^\top \Sigma^{-1} B^\top \lambda_2 \\ & - \quad \Big(\frac{1}{2} w_1 1^\top \Sigma^{-1} b + \frac{1}{2} w_2 1^\top \Sigma^{-1} r + 2P_D\Big)\lambda_1 \\ & - \quad \Big(\frac{1}{2} w_1 B \Sigma^{-1} b + \frac{1}{2} w_2 B \Sigma^{-1} r + 2c\Big)^\top \lambda_2. \end{aligned}$$

Where we have defined $\Sigma \doteq w_1 A + w_2 Q$. Note that Σ^{-1} exist and even more is a symmetric diagonal matrix if $w_1 a_{ii} + w_2 q_{ii} \neq 0$, $\forall i = 1, \ldots, N$; condition that holds true due to the entries in the EED problem.
Then we must finally solve,

$$\begin{aligned} \min_{\substack{\lambda_1 \in \mathbb{R} \\ \lambda_2 \in \mathbb{R}_+^{2N}}} \; & \Big\{\lambda_1^2 1^\top \Sigma^{-1} 1 + \lambda_2^\top B \Sigma^{-1} B^\top \lambda_2 + 2\lambda_1 1^\top \Sigma^{-1} B^\top \lambda_2 + \Big(\frac{1}{2} w_1 1^\top \Sigma^{-1} b + \frac{1}{2} w_2 1^\top \Sigma^{-1} r + 2P_D\Big)\lambda_1 \\ & + \Big(\frac{1}{2} w_1 B \Sigma^{-1} b + \frac{1}{2} w_2 B \Sigma^{-1} r + 2c\Big)^\top \lambda_2\Big\} \end{aligned}$$

which become a relatively simple Quadratic Programming problem, because there only exist constraints of non-negativity over λ_2. In matrix notation:

$$\min_{\substack{\bar{\lambda}_i \geq 0 \\ i \in \{2, \ldots, 2N+1\}}} \; \frac{1}{2} \bar{\lambda}^T \bar{A} \bar{\lambda} + \bar{b}^T \bar{\lambda}, \tag{1.11}$$

where:

$$\bar{\lambda} = \begin{pmatrix} \lambda_1 \\ \lambda_2 \end{pmatrix} \ , \quad \bar{A} = \begin{pmatrix} 1^\top \Sigma^{-1} 1 & 1^\top \Sigma^{-1} B^\top \\ B\Sigma^{-1} 1 & B\Sigma^{-1} B^\top \end{pmatrix},$$

$$\bar{b} = \begin{pmatrix} \frac{1}{2} w_1 1^\top \Sigma^{-1} b + \frac{1}{2} w_2 1^\top \Sigma^{-1} r + 2P_D \\ \frac{1}{2} w_1 B\Sigma^{-1} b + \frac{1}{2} w_2 B\Sigma^{-1} r + 2c \end{pmatrix},$$

and necessarily $0 \preccurlyeq \bar{A} \in \mathbb{R}^{(2N+1)\times(2N+1)}$ for problem (QP) have a solution. The next proposition prove that indeed, matrix $\bar{A}$ is positive semidefinite.

Proposition 1.5. *Let matrix $\bar{A} \in \mathbb{R}^{(2N+1)\times(2N+1)}$ be as above. Then for all vector $x \in \mathbb{R}^{2N+1}$ we have that:*

$$x^\top \bar{A} x = \sum_{i=1}^{N} \sigma_{ii} (x_{i+1} - x_1 - x_{i+1+N})^2 \geq 0, \tag{1.12}$$

where we have defined $\Sigma^{-1} \doteq (\sigma_{ii})$.

Proof. We make the proof of the proposition by induction over N. It can be easily prove that, for $N = 1$, (1.12) holds true. Suppose now that also holds true for $N = n \in \mathbb{N}$ and we are going to prove that it remains true for $N = n + 1$.
First note that if $N = n$, then:

$$\bar{A} = \begin{pmatrix} 1^\top \Sigma^{-1} 1 & -1^\top \Sigma^{-1} & 1^\top \Sigma^{-1} \\ -\Sigma^{-1} 1 & \Sigma^{-1} & -\Sigma^{-1} \\ \Sigma^{-1} 1 & -\Sigma^{-1} & \Sigma^{-1} \end{pmatrix}_{(2n+1)\times(2n+1)}$$

and for any $\mathbb{R}^{2n+1} \ni x = (x_1, x_2, x_3)^\top$, $x_1 \in \mathbb{R}$, $x_2, x_3 \in \mathbb{R}^n$, we have:

$$x^\top A x = x_1^2 1^\top \Sigma^{-1} 1 - 2x_1 x_2^\top \Sigma^{-1} 1 + 2x_1 x_3^\top \Sigma^{-1} 1 - 2x_2^T \Sigma^{-1} x_3 + x_2^T \Sigma^{-1} x_2 + x_3^T \Sigma^{-1} x_3. \tag{1.13}$$

Now, suppose that $N = n + 1$, then we have the matrix:

$$\bar{A} = \begin{pmatrix} 1^\top \Sigma^{-1} 1 + \sigma_{n+1} & -1^\top \Sigma^{-1} & -\sigma_{n+1} & 1^\top \Sigma^{-1} & \sigma_{n+1} \\ -\Sigma^{-1} 1 & \Sigma^{-1} & 0 & -\Sigma^{-1} & 0 \\ -\sigma_{n+1} & 0^\top & \sigma_{n+1} & 0^\top & -\sigma_{n+1} \\ \Sigma^{-1} 1 & -\Sigma^{-1} & 0 & \Sigma^{-1} & 0 \\ \sigma_{n+1} & 0^\top & -\sigma_{n+1} & 0^\top & \sigma_{n+1} \end{pmatrix}_{(2n+3)\times(2n+3)} .$$

Where we have defined $\Sigma^{-1} \doteq (\sigma_{ii})$ and 0 the null vector of $\mathbb{R}^n$. For simplicity in the notation also we made $\sigma_{n+1} \doteq \sigma_{n+1,n+1}$. Now we make $x^\top = (x_1, x_2^\top, x_{n+2}, x_3^T, x_{2n+3})$, then:

$$\begin{aligned} x^\top \bar{A} x \ &= \ x_1^2 (1^\top \Sigma^{-1} 1 + \sigma_{n+1}) - 2x_1 x_2^\top \Sigma^{-1} 1 + \ 2x_1 1^\top \Sigma^{-1} x_3 - 2x_2^\top \Sigma^{-1} x_3 + x_2^\top \Sigma^{-1} x_2 \\ &+ \ x_3^\top \Sigma^{-1} x_3 - 2x_1 x_{n+2} \sigma_{n+1} + 2x_1 x_{2n+3} \sigma_{n+1} 2x_{n+2} x_{2n+3} \sigma_{n+1} + x_{2n+3}^2 \sigma_{n+1} + x_{n+2}^2 \sigma_{n+1}. \end{aligned}$$

Completing square on x_{n+2} we obtain:

$$\begin{aligned}\sigma_{n+1}(x_{n+2}^2 - 2x_1x_{n+2} - 2x_{n+2}x_{2n+3}) &= \sigma_{n+1}(x_{n+2} - x_1 - x_{2n+3})^2 - \sigma_{n+1}x_1^2 - \sigma_{n+1}x_{2n+3}^2 \\ &- 2\sigma_{n+1}x_1x_{2n+3}\end{aligned}$$

Therefore, if $N = n + 1$ we have:

$$x^\top \bar{A}x = \sum_{i=1}^{n} \sigma_{ii}(x_{i+1} - x_1 - x_{i+1+n+1})^2 + \ \sigma_{n+1,n+1}(x_{n+2} - x_1 - x_{2n+3})^2$$

where we have used that assumption of (1.12) holds true for n and (1.13). Then, note that:

$$\sigma_{n+1,n+1}(x_{n+2} - x_1 - x_{2n+3})^2 = \sigma_{ii}(x_{i+1} - x_1 - x_{i+1+n})^2,$$

with $i = n + 1$ and $n = n + 1$. □

Therefore $\nu \in \mathbb{R}$ and we can use any efficient algorithm for Quadratic Programming to find its value; and the optimal value of the primal problem can be recovered thanks to the Strong Duality property, which ensure that $\nu = \mu$.
We point out that the proposed approach could be applied for problems where power losses are considered as a fraction of the power demand or as a lineal function of the output power of the generators. When power losses are considered as a quadratic function of the power output, an explicit equation as (1.10) becomes more involved, but instead, a linearization strategy could be used.
In the next section we study three problems, showing that our approach provides better solutions than those obtained by other authors.

1.4 Study-cases

In order to show some advantages of our duality approach, we discuss three different problems with data obtained from [69, 142, 165]. The solution to all the problems were obtained through MATLAB on a computer with Intel(R) Core(TM) i7-4790S processor (3.20 GHz × 8) and 8 GB of memory. The principal motivation for the numerical test was to check the general performance of the proposed method. Hence, as a first approach, problem (1.11) was solved by using the built-in MATLAB routine for Quadratic Programming convex problems (*quadprog*) and selecting, randomly, a high number of solutions (1000) on the Pareto front, i.e., the weight vector is randomly selected.

1.4.1 A 6-generators system

In the first case we analyze the performance on a test system of 6 generating units [69], with a power demand of 1000 MW and characteristics in Table 1.1.
In [69] the authors used a gravitational search algorithm (GSA) to solve the EED problem and compare their results with those obtained in previous works. They find what they call "the best EED results" with the schedule shown in Table 1.2.
We point out some facts from the solution obtained by our method, when the proposed duality approach is applied, see Table 1.2:

Table 1.1: 6-generating units system data.

Unit i	c_i	b_i	a_i	s_i	r_i	q_i	$p_{\min_i}$	$p_{\max_i}$
1	756.800	38.540	0.1525	13.860	0.3300	0.0042	10	125
2	451.325	46.160	0.1060	13.860	0.3300	0.0042	10	150
3	1050.000	40.400	0.0280	40.267	-0.5455	0.0068	35	225
4	1243.530	38.310	0.0355	40.267	-0.5455	0.0068	35	210
5	1658.570	36.328	0.0211	42.900	-0.5112	0.0046	130	325
6	1356.660	38.270	0.0180	42.900	-0.5112	0.0046	125	315

Table 1.2: 6-generating units system results obtained by GSA and the proposed approach.

Element	GSA	Duality
Power Output Generator 1	78.821	80.9197
Power Output Generator 2	83.0013	80.6684
Power Output Generator 3	164.2907	165.4529
Power Output Generator 4	164.9136	164.2878
Power Output Generator 5	258.1108	255.1709
Power Output Generator 6	250.8619	253.5003
Cost ($/h)	51255.7880	51252.24827
Emissions (kg/h)	827.1380	827.1104821

Table 1.3: 6-generating units system results comparison.

Method	Cost ($/h)	Emission (kg/h)
γ-iteration [12]	51264.6	828.720
Recursive [12]	51264.5	828.715
PSO [12]	51269.6	828.863
DE [12]	51264.6	828.715
Simplified recursive [12]	51264.6	828.715
GA similarity [70]	51262.31	827.261
Proposed GSA [69]	51255.7880	827.1380
Duality Approach	51252.24827	827.1104821

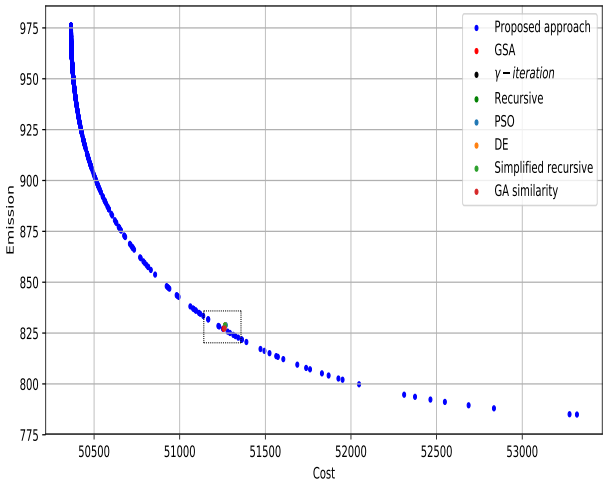


Figure 1.1: Pareto front generated by the method (6 generating units, 1000 non-dominated solutions).

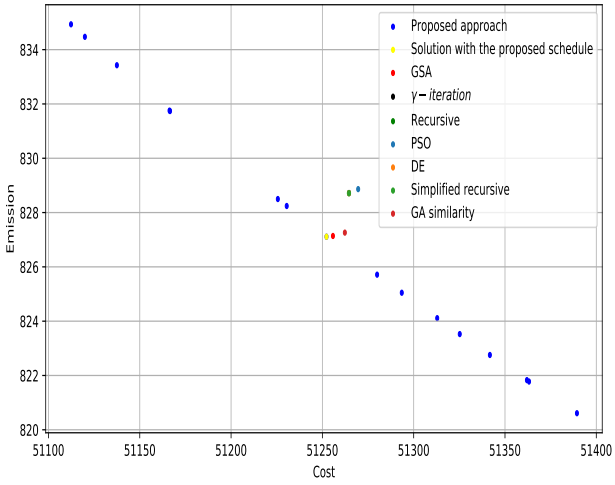


Figure 1.2: A zoom at the Pareto front generated by the method (6 generating units, 1000 non-dominated solutions).

- As expected, the entire Pareto front can be obtained by varying the weights and consequently a trade-off curve for the fuel cost and the amount of emission can be estimated, see Figure 1.1.
- Clearly, solutions on the Pareto front can not be compared; but respect to other investigations we were able to improve some previously reported values with the specific proposed schedule, see Table 1.3. Note that we found solutions with simultaneously cost and emission lowest than those obtained in [12, 69, 70], see Figure1.2.
- Although compilation time in [69] was not presented, we want to highlight that we were able to find the solutions in approximately 5 seconds and finding 1000 points on the Pareto front.
- Another important aspect is that no violation of the total power demand constraint occurs in our schedule, we get $\sum_{i=1}^{6} p_i = 1000$.

1.4.2 A 11-generators system

Here we consider the 69-bus, 11 generators system. The total power demand is $P_D = 2500$ MW and the system characteristics can be seen on Table 1.4. With the proposed approach solutions were obtained and compared with the results presented in [165], where several metaheuristics are tested, see Table 1.5. In this table appears the solution obtained for the EED problem selecting one specific weight vector and NA means that this information was not available.

Table 1.4: 11-generating units system data.

Unit i	c_i	b_i	a_i	s_i	r_i	q_i	$p_{\min_i}$	$p_{\max_i}$
1	387.85	1.92699	0.00762	33.93	-0.67767	0.00419	20.0	250.0
2	441.62	2.11969	0.00838	24.62	-0.69044	0.00461	20.0	210.0
3	422.57	2.19196	0.00523	33.93	-0.67767	0.00419	20.0	250.0
4	552.50	2.01983	0.00140	27.14	-0.54551	0.00683	60.0	300.0
5	557.75	2.22181	0.00154	24.15	-0.40006	0.00751	20.0	210.0
6	562.18	1.91528	0.00177	27.14	-0.54551	0.00683	60.0	300.0
7	568.39	2.10681	0.00195	24.15	-0.40006	0.00751	20.0	215.0
8	682.93	1.99138	0.00106	30.45	-0.5116	0.00355	100.0	455.0
9	741.22	1.99802	0.00117	25.59	-0.56228	0.00417	100.0	455.0
10	617.83	2.12352	0.00089	30.45	-0.41116	0.00355	110.0	460.0
11	674.61	2.10487	0.00098	25.59	-0.56228	0.00417	110.0	465.0

In this case we point out some facts from the comparison of the results obtained by our method with those discussed in [165]:

- With the proposed approach, several existing results were improved [12, 69, 70]. This is evident when we look at and compare the Pareto front obtained, see Figure 1.3. A closer look at the Pareto front shows how the results of [165] (that were not directly comparable with the reported in [69]) are also dominated by the solutions obtained with the proposed

approach; particularly, we show a specific schedule in Table 1.5 for a better understanding of this statement.

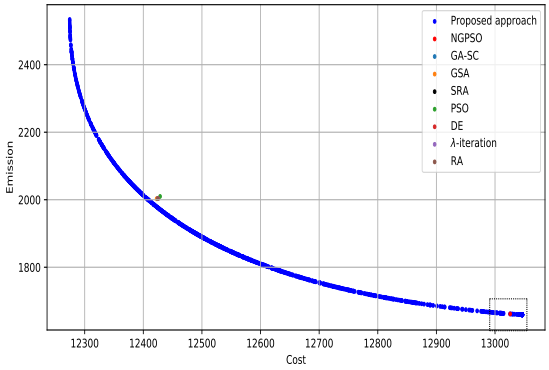


Figure 1.3: Pareto front generated by the method (11 generating units, 1000 non-dominated solutions).

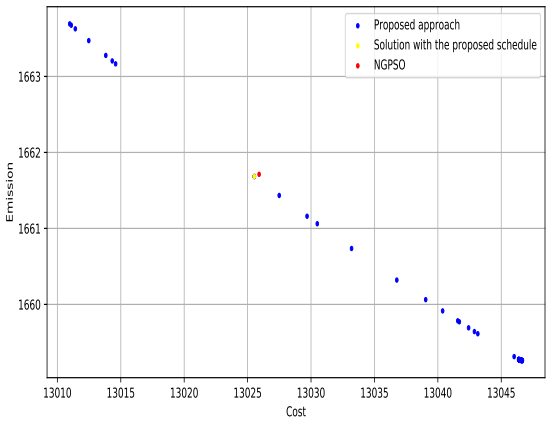


Figure 1.4: A zoom at the Pareto front generated by the method (11 generating units, 1000 non-dominated solutions).

- When we generate 1000 non-dominated solutions compilation time was around 7 seconds, Figure 1.3; and our approach generated what clearly appears to be a comprehensive Pareto front. But when we just search 50 non-dominated solutions compilation time now oscillates around 0.5 seconds and a well-distributed (smooth and uniform) Pareto front still can be obtained, see Figure 1.5.

Table 1.5: 11-generating units system results comparison.

Unit i	GA-SC [70]	GSA [69]	SRA [12]	PSO [12]	DE [12]	λ-iteration [12]	RA [12]	NGPSO [165]	Duality
1	138.8618	138.9382	139.672	NA	NA	NA	NA	243.3349	243.2094
2	112.1312	110.2728	112.781	NA	NA	NA	NA	210.0	210.0
3	146.7169	147.9728	145.802	NA	NA	NA	NA	250.0	250.0
4	222.1041	221.1072	221.527	NA	NA	NA	NA	169.0338	169.0471
5	137.1962	137.7986	136.774	NA	NA	NA	NA	142.6156	142.6563
6	217.3208	217.9015	218.578	NA	NA	NA	NA	168.8431	168.8602
7	140.4711	141.3801	140.261	NA	NA	NA	NA	142.5922	142.6051
8	348.9008	349.6497	345.46	NA	NA	NA	NA	317.2895	317.3080
9	326.5188	327.3178	329.484	NA	NA	NA	NA	276.5437	276.5512
10	363.5275	363.4766	363.645	NA	NA	NA	NA	303.2289	303.2378
11	346.2508	344.1847	346.430	NA	NA	NA	NA	276.5181	276.5249
Cost	12423.77	12422.66	12424.94	12428.63	12425.06	12424.94	12424.94	13025.9072	13025.5290
Emission	2003.0304	2002.9499	2003.3000	2009.7200	2003.3500	2003.3010	2003.3000	1661.7118	1661.6833

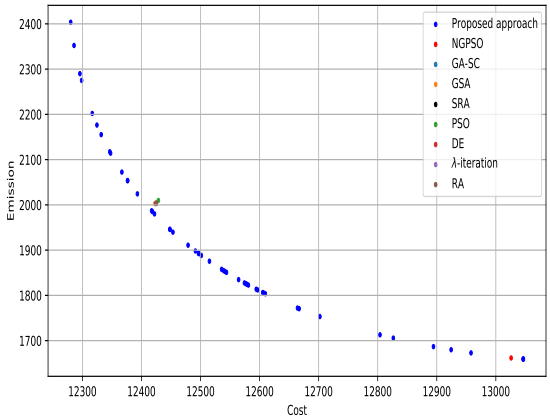


Figure 1.5: Pareto front generated by the method (11 generating units, 50 non-dominated solutions).

- There is not power balance constraint violation with the proposed schedule. A result that is similar to the obtained in [69], but better than the obtained in [165]; where $\sum_{i=1}^{6} p_i = 2502.9998$.

1.4.3 A 40-generators system

In this case a 40 generating units system is analyzed. The system data can viewed in [142] and the power demand $P_D = 10500$ MW. In [142] the authors proposed a Predictor-Corrector Primal-Dual Interior Point Method and solve the problem using WSM and the ϵ-Constrained Method obtaining the results showed in Figure 1.6. From the application of the proposed duality approach for the solution of this system we remark that:

- Compilation time was around 8 seconds. This is a reasonable time considering that 1000 solutions were found over $\mathbb{R}^{40}$. Even more, this compilation time remains stable, i.e., do not depend of the selection of an initial parameter set-up, as in [142].
- We improve the results of [142]. This is showed in Figure 1.6, where solutions found by us dominate the previously obtained.

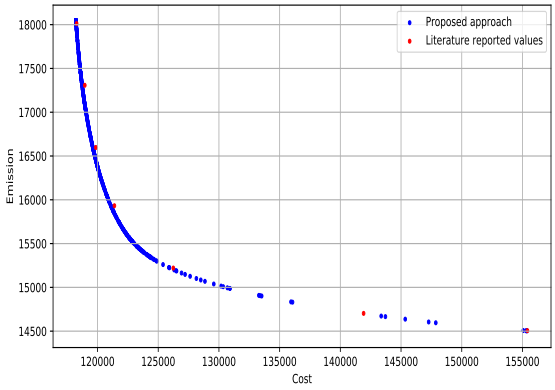


Figure 1.6: Comparison between the proposed method and previous one for the 40 generating units system.

1.5 Improving Pareto fronts' shape

As it is known, a main drawback of the WSM is the criterion used to weights selection. Analyzing the previous study-cases we can observe that, a random selection of the weights allowed us to obtain the corresponding Pareto fronts, with low computational times. However, these solutions on the Pareto fronts are not well distributed, showing a high density in certain regions (look at the upper-left corner of Figure 1.6 or Figure 1.1); thus, providing the DM no useful information about the trade-off between cost and emission elsewhere. As alternative, we propose a simple heuristic

strategy, based on bisection algorithm, able to find solutions on the Pareto front preserving the Euclidean distance between each of them. The main idea, see the pseudocode for the algorithm 1, is to obtain the ending points of the Pareto front and look for new solutions, while weights are modified, that keep distance between consecutive solutions as approximately equal. For simplicity, the distance between the last two solutions is set as free.

Data: $0 = \mathbf{w_0} \longleftrightarrow \mathbf{P_0}$, $1 = \mathbf{w_N} \longleftrightarrow \mathbf{P_N}$,
Result: $\mathbf{w_i} \longleftrightarrow \mathbf{P_i}$, $i = 0, ..., N$.

1 Define the approximately Euclidean distance ($\mathbf{D}$) to keep between consecutive solutions, except $\mathbf{P_N}$ and $\mathbf{P_{N-1}}$;

2 Calculate $\mathbf{P_1}$ such as $d(P_1, P_0) \approx D$, by an iterative "bisection" strategy with

$$w_1^j = \frac{w_0^j + w_N^j}{2}$$

3 **i**=1; **while** $d(P_i, P_N) \not\approx D$ **do**

4 $\quad \mathbf{w_0} = \mathbf{w_i} \longrightarrow \mathbf{P_0} = \mathbf{P_i}$;

5 $\quad$ Search for $\mathbf{P_{i+1}}$ such as $d(P_{i+1}, P_i) \approx D$ by an iterative "bisection" strategy with

$$w_{i+1}^j = \frac{w_0^j + w_N^j}{2}$$

$\quad \mathbf{i} = i + 1$;

6 **end**

Algorithm 1: Finding weights that keep Euclidean distance between consecutive solutions approximately as constant.

To evaluate the results obtained with this strategy, we compute an indicator measuring the diversity of the solutions. Particularly, we use the *normalized diversity metric* (DMn), calculated as follows [71]:

$$DMn = \frac{d_f + d_l + \sum_{i=1}^{S} |d_i - \bar{d}|}{d_f + d_l + (S-1)\bar{d}},$$

where d_i is the Euclidean distance between consecutive solutions in the obtained non-dominated set of solutions and $\bar{d}$ is the average of all distances d_i ($i = 1, .., S$), assuming there are S solutions in the obtained non-dominated set. The parameters d_f and d_l are the Euclidean distances between the extreme and the boundary solutions. This metric gives smaller values to better distributions and the most widely and uniformly spreadout set of non-dominated solutions returns a DMn of zero.
By using the proposed strategy, in the case of the 6-generators system we obtained the Pareto front in Figure 1.7. In this case the DMn drops from 1.71 with the random strategy to just 0.005. Note

that this value is even closest to zero than those obtained in [38, 45], where bi-objective problems are also studied. If we compare Figure 1.1 and 1.7, it is clear the improvement obtained. Note that solutions are now well distributed over the entire Pareto front, even in the lower-right corner, filling the previously existing gaps.

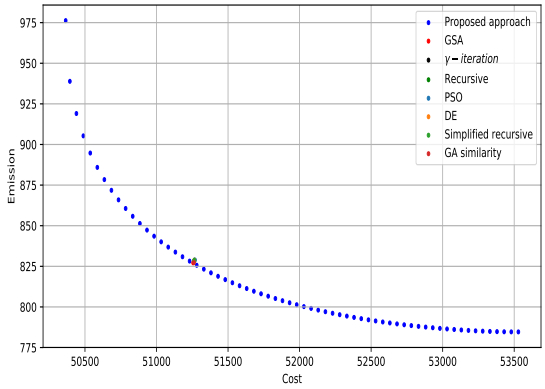


Figure 1.7: Pareto front for the 6-generators system based on the heuristic strategy (1000 non-dominated solutions).

On another hand, for the 11-generators system we can observe in Figure 1.8 the new Pareto front obtained. We can see now a better distribution, even just searching for 50 non-dominated solutions. In this case the DMn also decreased, from 0.57 to 0.02.

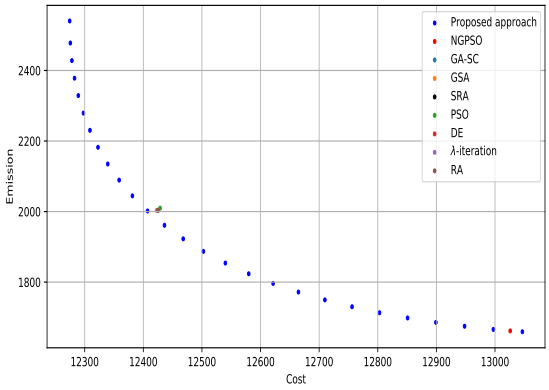


Figure 1.8: Pareto front for the 11-generators system based on the heuristic strategy (50 non-dominated solutions).

Finally, for the 40-generators system we obtained now the Pareto front in Figure 1.9. In this case the DMn drops from 0.96 to 0.05, improving also the shape of the Pareto front obtained in [142], where the $DMn = 0.87$.

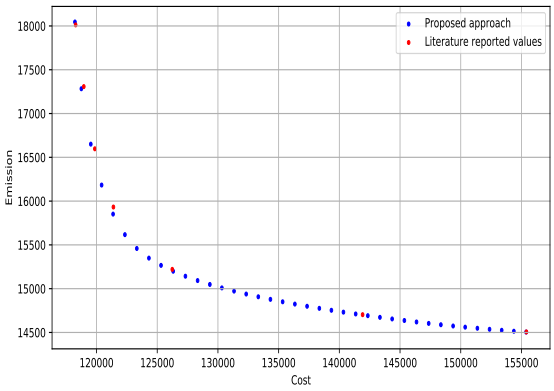


Figure 1.9: Pareto front for the 40-generators system based on the heuristic strategy.

1.6 A final note about the obtained non-dominated solutions

Finally we point out that, under strong assumptions, analytical solutions can be obtained for the EED problem. Althought in general, these exact solutions are difficult to obtain, particularly for the 6-generators system and the 11-generators system discussed above, analytical solutions were obtained. This allowed us to assess the performance of the solutions obtained by the duality theory approach proposed. This is the main result in [31]; whereas the mathematical development of the duality theory approach, is the main discussion in [32]. In Figures 1.10,1.11 and 1.12 we can observe that solutions obtained by the proposed approach are closest to the corresponding analytical solutions, than those obtained by metaheuristic algorithms. Therefore, we indeed have outperformed the previous results.

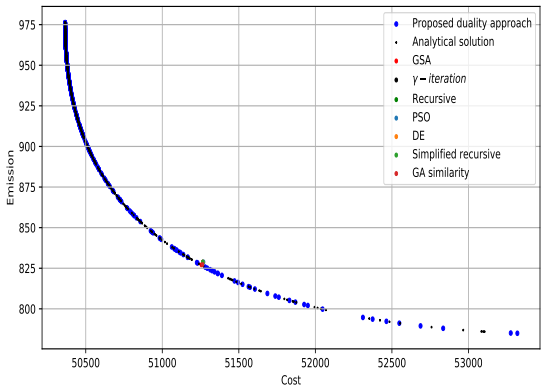


Figure 1.10: Comparison between analytical and duality theory based solutions (6-generators system).

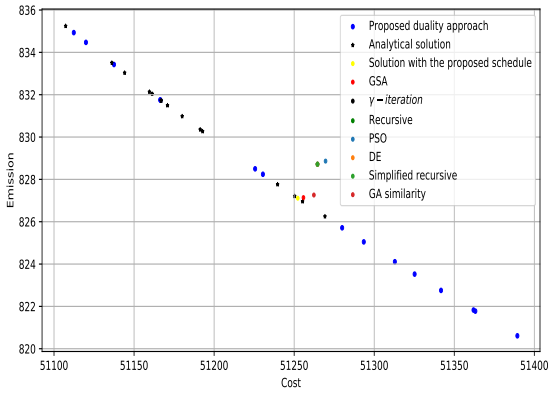


Figure 1.11: Comparison between analytical and duality theory based solutions (zoom at the 6-generators system).

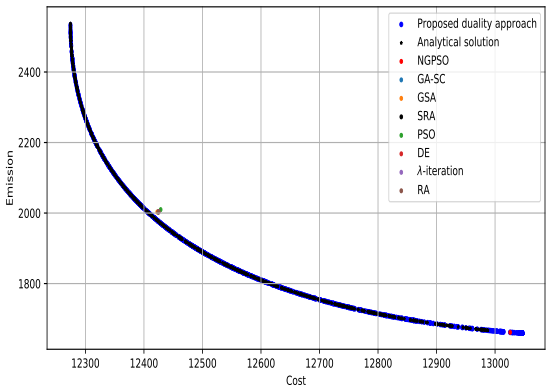


Figure 1.12: Comparison between analytical and duality theory based solutions (11-generators system).

1.7 Conclusions and future works

In general, the proposed method performs well in all the studied cases, where clearly several existing results were improved. The analytical nature of our approach makes not initial parameter set-up needed and that solutions obtained in each run be feasible, therefore is a robust method.
However, during the develop of this investigation, some evident aspects that could be improved were detected and it motivate future works. A more efficient selection of weight vector could be done. Note that the heuristic algorithm proposed to improve the Pareto fronts, needs multiple runs. Therefore, with other strategies we could diminish even more the computational time. It must be analyzed the implementation of better suited algorithms for solving the Quadratic Programming problem (1.11). This will be beneficial in bigger power networks, since computational time can be diminished. Also it must be included power losses, but note that when power loss is considered as a fraction of the power demand or as a linear function of the power output, our method is straightforward. Finally, a strategy to point out a single solution from the Pareto front is also needed, since often the decision-maker need just one solution. An alternative could be to considerate the non-dominated solution with minimum power loss.

Chapter 2

Models uncertainties effect on the ECED

2.1 Introduction

Although electricity is a clean and relatively safe form of energy when it is used, the generation and transmission of electricity have severe effects on the environment; particularly when thermal power plants (burning coal, natural gas or oil), are used to the generation. According to [68], power generation with fossil fuels comes with significant costs to the environment and human health. Combustion releases emissions of sulfur dioxide (SO_2), nitrogen oxides (NO_x), particulate matter (PM), carbon monoxide (CO), volatile organic compounds (VOCs) and various trace metals like mercury, into the air through stacks that can disperse this pollution over large areas. Environmental effects of the above mentioned pollutants include impaired visibility, damage to materials, damage to vegetation, deposition as acid rain, ozone formation and contribution to the greenhouse effect [100]. On another hand, in Table 2.1, major air pollutants and their associated health hazards are presented.

Table 2.1: Pollutants and their hazards [159].

Name of Pollutant	Health Impacts
Respirable PM	Respiratory illness, including chronic bronchitis and asthma; heart diseases
SO_2	Heart diseases; respiratory problems including pulmonary emphysema, cancer, eye burning, headache
NO_2	Lung irritation, viral infection, airway resistance, chest tightness
Suspended PM	Pneumoconiosis, restrictive lung diseases, asthma, cancer
Ozone	Impaired lung function, chest pain, coughing, irritation of eyes and nose
CO	Cherry lips, unconsciousness, death by asphyxiation

Therefore, emission problems corresponding to the fossil fuels-based power plants cannot be neglected. According to [95], many strategies like installation of air filters, pollutant-cleaning equipment, switching to low emission fuels, replacement of aged equipment and generating units and integration of renewable energy sources, have taken into consideration in order to minimize the emissions. But all of them need considerable amount of capital and huge replacement/modification or upgrade strategies that can be considered as long-term solutions. Thus, emission dispatch strategies become attractive solutions, since are easy to implement and require less additional costs [50, 143].

The incorporation of environmental concerns to the usual economic dispatch have been treated by authors in different manners. On the Environmental/Economic Dispatch (EED) problem, authors treat the polluting emissions and the fuel cost as two conflicting objectives which are optimized simultaneously subjected to the practical constraints [30,123]. Whereas in the Emission Constrained Economic Dispatch (ECED) problem, fuel cost is minimized while treating emissions as a constraint with a pre-specified limit [2,138,156].

The ECED problem is closely related to emission norms adopted in several regions and established in order to improve local or regional air quality index (AQI), see for example [144]; air quality that is affected by the emissions dispersed in the atmosphere and also with weather conditions (temperature, precipitation and specially wind speed and direction). Note that, although environmental standards restrict emissions levels from most thermal power plants, implementation of the same norm across a country will not ensure uniform air quality. Therefore, a more effectively strategy could be to implement emission standards for thermal power plants region-wise, in order that power plants located at critically polluted areas will have stricter (lower) emissions limits than those located at areas with good AQI's. In such cases, the ECED provides a strategy to manage the pollutant emissions from particular sets of generating units; obtaining optimal solutions within the maximum emission level specified.

As was pointed out in [48], the final accuracy of the optimization process depends on many factors, such as the selected models, the data and parameters used as well as the technique or the tool used for solving the formulated problem. Several methods are generally used for establishing the fuel cost and emission functions of generating units: performance testing, determination from operating records or use of manufacturer's guarantee data adjusted to actual operating conditions [78]. But, high cost of performance testing and possible incorrect representation of the shape of the input-output curves provided by manufacturer's data, makes the determination from operating records a suitable alternative; even more if we consider that, nowadays, operating data is readily available in many generation companies.

By analyzing the literature, we can see that research efforts have been continually carried regarding to model selection and estimation of its parameters, as well as the tools to obtain these parameters from operating records. Least square (LS) is used in [64] to on-line steam unit dispatch, resulting in minimum emissions. Several tests were carried out to represent the input-output (I/O) curves and Newton-Raphson was used to overcome the mathematical difficulties derived from the non-polynomial model selected. In [47], four algorithms for estimation of the parameters of models used in optimal economic operation of electric power systems, are proposed: Weighted Least Squares, Gauss-Newton Method or Bard-Algorithm, Marquardt Algorithm and Powell Regression Algorithm. A second order function is taken as the fuel cost curve model. In [34] authors present an on-line input-output curve identification based on the Linear Sequential Regression Technique and Weighted Least Square Linear Regression. A method is proposed to estimate the actual fuel cost function of power stations in [49]. The coefficients are obtained by using operating records and LS regression technique. To obtain the representation of generators' fuel costs, two polynomial curve fitting methods, Gram-Schmidt orthonormalization and LS, are evaluated in [90]. In [136] an algorithm is proposed, based on Least Absolute Value approximations, for estimating the coefficients of a fuel cost curve, used for economic dispatch. The Kalman-Filtering algorithm for the on-line identification of the I/O curve of thermal power plant is used in [135]. In [15] an environmental dispatch algorithm is presented. In here the minimization of SO_2 and NO_x emissions is addressed by considering quadratic models and parameters computed via the LS criteria. Several models

of input-output curves of thermal units are considered in [7]. A Genetic Algorithm (GA) is used and the accuracy of the results is evaluated by analyzing the difference between the actual cost and the estimated cost, once the parameters were fitted. In [9, 48] a method for estimating the parameters of fuel cost functions is proposed and tested. The total estimation error such that the selected model follows field data measurements as closely as possible is minimized by using a Particle Swarm Optimization (PSO) algorithm. The authors obtained more accuracy compared to when LS technique was used. In [139] an Artificial Bee Colony (ABC) algorithm is used to estimate the fuel cost curve parameters of thermal power plants. The total error with the estimated parameters is the performance index used to compare the results obtained with the ABC algorithm and those obtained with GA, PSO and LS. In [43, 44] a Teaching Learning Based Optimization is used in the economic dispatch and results are compared with ABC, PSO and LS. In [125] a Differential Evolution (DE) algorithm is proposed for estimating the optimal parameters of fuel cost curves; whereas the Modified Radial Movement Optimization technique is used in [147]. In [10] a Crow Search Algorithm is proposed for accurate estimation of the input-output characteristics of thermal power plants with and with-out valve-point effect. In [126] the authors proposed an Improved Differential Evolution (IDE)-based technique to estimate the optimal parameters of thermal power plants. A Modified Whale Optimization (MWO) algorithm and a Chaotic MWO algorithm are used to solving the problem of parameters estimation for input-output curves of thermal and hydro generating units in [128]. In [140] the Improved Symbiotic Organism Search algorithm is proposed to estimate parameters of smooth and non-smooth fuel cost functions for improving the solution accuracy of economic dispatch problems.

However, in none of the above papers, authors deal with the inaccuracies related to parameter estimations. In general, the coefficients obtained are considered as fixed values and the subsequent optimization problem is solved. This deterministic approach can lead to significantly errors in the proposed optimal schedule, specially when few operating points are used to fit the models. Even more, in the case of the ECED problem, uncertainties can lead to obtain infeasible environmental solutions, i.e., solutions with a high probability associated of generate emissions above the pre-established limit.

In this chapter we propose an approach that, based on the statistical intervals of the experimentally derived models, is able to deal with those uncertainties. This allow us, for the ECED problem, to obtain optimal solutions that remain environmentally feasible, or at least risk of emission limits' violations is diminished. The proposed strategy is tested in a real 4-generating units system from the Chilean electrical power network and the results obtained show that optimal solutions based on statistical intervals, can minimize the risk of environmental violations, by accepting a slightly increase in the fuel cost. As we show in the next section, this approach entirely differs from previous ones, where expected mean values are considered instead.

The remainder of this chapter is organized as follows. In Section 2.2 a review is conducted about previous investigations dealing with emission and fuel cost function's uncertainties. In Section 2.3 fuel cost and emission models are presented; multiple linear regression is used to estimate the respectively coefficients and the statistical intervals associated. The proposed approach for solving the Environmentally Constrained Economic Dispatch based on statistical intervals is developed in Section 2.4. A study-case based on four generating units from the Chilean electrical power network is analyzed in Section 2.5. Finally, conclusions and some extensions of this work are in Section 2.6.

2.2 Literature review

Alternatives to deal with non-deterministic coefficients, both for fuel cost and emission functions, have already been proposed by several authors. But, as we will see next, have been mainly focused on the expected values of such coefficients.

In [67], authors analyzed the performance of a two-generating units system, when incremental cost, used to solve the ED problem, is assumed incorrectly represented, as being higher for one generating unit and lower for the other; an analytical determination of the economical loss is obtained. A sensitivity analysis is used in [148] for correcting the economic dispatch when small order changes in the system input parameters and cost function coefficients variations are considered. A four plants system, with two thermal plants and two hydro plants is studied. At least in our knowledge, among the earliest papers dealing with coefficients uncertainties and combined economic-emission objectives are [40, 41], where the Economic Emission Load Dispatch (EELD) problem is studied. Considering uncertainties in the system production cost and nature of the load demand, a stochastic EELD problem is formulated. The fuel cost coefficients, emission coefficients and load demand are considered random variables and the stochastic model is converted to its deterministic equivalent by taking the expected values, with the assumption that all the random variables are normally distributed and statistically dependent on each other. The Weighted Sum Method (WSM) is used to solve the multiobjective problem, where expected fuel cost, NO_x emissions and deviations due to unsatisfied load are the objectives to minimize. The fuel cost curve is assumed to be approximated by a quadratic function of the generator output and the expected value is obtained by expanding the function, using Taylor's series, about the mean, is represented by:

$$\overline{F}_1 = \sum_{i=1}^{N} [\overline{a}_i \overline{P}_i^2 + \overline{b}_i \overline{P}_i + \overline{c}_i + \overline{a}_i \mathrm{var}(P_i) + 2\overline{P}_i \mathrm{cov}(a_i, P_i) + \mathrm{cov}(b_i, P_i)],$$

where $\overline{P}_i$ is the expected value of the generator output and $\overline{a}_i, \overline{b}_i$ and $\overline{c}_i$ are the expected cost coefficients. A similar expression is obtained for the expected NO_x emission:

$$\overline{F}_2 = \sum_{i=1}^{N} [\overline{d}_i \overline{P}_i^2 + \overline{e}_i \overline{P}_i + \overline{f}_i + \overline{d}_i \mathrm{var}(P_i) + 2\overline{P}_i \mathrm{cov}(d_i, P_i) + \mathrm{cov}(e_i, P_i)],$$

where $\overline{d}_i, \overline{e}_i$ and $\overline{f}_i$ are the expected emission coefficients. A similar treatment of uncertainties is realized in [13], but considering additionally objectives and using evolutionary and Hookes-Jeeves methods. In [4, 5] three cases are studied depending on the stochastic variables considered: power generated, power generated and system load, and power generated, system load and cost and emission coefficients. They compared the results against the deterministic non-dominated solutions, obtaining that stochastic solutions are generally dominated by the deterministic ones, meaning that in practice, real world operation cost and emission would be higher. The authors assumed decision variables P_{G_i} as normally distributed with mean P_{G_i} and standard deviation $\sigma_i = 0.1 P_{G_i}$. Simulations are performed for each solution having mean P_{G_i} and standard deviation are created within $2\sigma_i$. In the stochastic approach, the objective functions are reformulated as:

$$
\begin{aligned}
&\min \overline{\text{Cost}} + 2\sigma_{\text{Cost}} \\
&\min \overline{\text{NO}}_{\text{x}} + 2\sigma_{\text{NO}_{\text{x}}} \\
st \quad &\textstyle\sum_{i=1}^{n} P_{G_i} - P_D - P_L = 0 \\
&P_{G_i \min} \leq P_{G_i} \leq P_{G_i \max} \\
&R \geq R^{cr},
\end{aligned}
$$

where $\overline{\text{Cost}}, \overline{\text{NO}}_{\text{x}}, \sigma_{\text{Cost}}, \sigma_{\text{NO}_{\text{x}}}$ are the expected cost, expected NO_{x} emission and standard deviation of the expected fuel cost and the expected NO_{x} emission, respectively. Reliability R is calculated according to the number of cases for which P_1 is found to be within $2\sigma_1$. This follows because required reliability R^{cr} is 95.6% for which $P_r\{\mu_1 - 2\sigma_1 < P_1 < \mu_1 + 2\sigma_1\}$. In [14] a stochastic multiobjective line security constrained problem is formulated to minimize the expected operating cost, polluting gas emission values and variance of active and reactive power generation, with explicit recognition of statistical uncertainties in thermal power generation cost coefficients, power demand and hence, power generation. The WSM is used to simulate the trade-off relationship between conflicting objectives. Assumptions, expected values and deviations are considered as in [40, 41]. In [149], a PSO algorithm is used to solve a stochastic economic emission dispatch problem. As in [40, 41], the stochastic model is obtained through Taylor's series expansion about the mean and the expected values of cost, emissions and power loss are the objectives to minimize. However, as can be noted, these relatively few investigations are mainly concerned on the uncertainties and optimization of expected mean response variables; which can lead to inaccurate results in some situations. As pointed out in [110], when coefficients of the fuel cost and emission functions are estimated via regression, the variance of the predicted values (for a specific power output) is bigger than for the expected value. This can lead to severe inaccuracies or even to obtain infeasible "optimal" solutions. Take for instance polluted areas, where depending on air quality, emission limits are established. Then, for a particular period, when the ECED problem is solved, we want to obtain an optimal solution that satisfy the emission level, not considering the mean emissions but the predicted point emission instead. The proposed approach developed in this chapter could be particularly beneficial in the Chilean context, where geographical location (Andes mountain range to the east and smaller coastal mountains in the west) and meteorological patterns of several cities, impede the ventilation and dispersion of air pollutants. This problem and its health consequences becomes specially acute in the April-August period [132]. Therefore, during these months, controlling the predicted emission level is more important than just controlling the mean emission level expected.

2.3 Experimentally derived fuel cost and emission functions

2.3.1 Fuel cost and emission models

As was pointed out in [43, 48], parameter estimation is a crucial issue in power systems operation. Several mathematical models are usually used for fuel cost and emissions representation, in this investigation we consider smooth functions, particularly polynomials [126, 128, 140]:

$$f_i(p_i) = a_{0i} + \sum_{j=1}^{M} a_{ji} p_i^j,\ i = 1, \ldots, N,$$
$$e_i(p_i) = \alpha_{0i} + \sum_{j=1}^{M} \alpha_{ji} p_i^j,\ i = 1, \ldots, N,$$

with $M = \{1, 2, 3\}$, N is the number of generating units and $f_i : \mathbb{R} \to \mathbb{R}$, $e_i : \mathbb{R} \to \mathbb{R}$ are the fuel cost and emission function of the ith generating unit, respectively.
To estimate the respectively coefficients, as was discussed in Section 2.1, different approaches have been used. We consider Simple Linear and/or Multiple Linear Regression to obtain coefficients and assess uncertainties associated to these coefficients, via its statistical intervals. In general, simple regression is based on the assumption that the true relationship between a quantitative response y and a single predictor variable x takes the form:

$$y = f(x) + \epsilon, \tag{2.1}$$

for some unknown function f and ϵ is a mean-zero random error term which is normally distributed. If f is to be approximated by a linear function, then we are in the Linear Regression case, and we can write relationship (2.1) as:

$$y = \beta_0 + \beta_1 x + \epsilon. \tag{2.2}$$

Linear regression (2.2) can be considered a particular case of the polynomial regression model [93]:

$$Y = X\beta + \epsilon,$$

where $Y \doteq (y_1, \ldots, y_n)^\top$ is a vector of observations; X is a $n \times (M + 1)$ full column-rank design matrix with the ith $(i = 1, \ldots, n)$ row given by $(1, x_i, \ldots, x_i^M)$; $\beta \doteq (\beta_0, \ldots, \beta_M)^\top$ is a vector of unknown coefficients and $\epsilon \doteq (\epsilon_1, \ldots, \epsilon_n)^\top$ is a vector of independent random errors with each $\epsilon_i \sim N(0, \sigma^2)$, where σ^2 is an unknown parameter.
In practice, coefficients β are unknown and must be estimated from observation pairs $(x_i, y_i), i = 1, \ldots, n$; in such a form that the model fits available data well. There are a number of ways measuring closeness (for the ECED problem see for example other criteria used in [43, 48]), but the most common approach is the Least Square criterion, which we used in here. It is well known that using LS criteria and solving the convex unconstrained problem associated, we can obtain the estimate of β as:

$$\hat{\beta} = (X^\top X)^{-1} X^\top Y. \tag{2.3}$$

2.3.2 Statistical intervals

However, (2.3) give us point estimates for the regression coefficients, but usually, a more useful result is to obtain a measure of the precision of the above estimates.
Confidence intervals (CI's), gives us a range of values (with a prescribed confidence level), that will cover the actual unknown population parameter that is being estimated from sample. For simple linear regression, a $100(1 - \alpha)$ percent CI for $E(y|x)$, the expected value of y for a given x, is [104]:

$$\hat{\beta}_0 + \hat{\beta}_1 x \pm t_{\alpha/2,n-2}\sqrt{\mathrm{MSR}\left(\frac{1}{\mathrm{n}} + \frac{(\mathrm{x}-\bar{\mathrm{x}})^2}{\sum_{\mathrm{i}=1}^{\mathrm{n}}(\mathrm{x_i}-\bar{\mathrm{x}})^2}\right)},$$

where $t_{\alpha/2,n-2}$ is the percentile from Student's t distribution with parameter equal to the degrees of freedom of the error and α equal to the acceptable degree of risk; n is the number of observation pairs used for the regression; x_i is the predictor value of the observation pairs; $\bar{x}$ is the average value of such predictors and the quantity MSR is called the residual mean square, and is obtained from:

$$\mathrm{MSR} = \frac{\sum_{\mathrm{i}=1}^{\mathrm{n}}(\mathrm{y_i}-\bar{\mathrm{y}}_\mathrm{i})^2 - \hat{\beta}_1 \sum_{\mathrm{i}=1}^{\mathrm{n}} \mathrm{y_i}(\mathrm{x_i}-\bar{\mathrm{x}})}{\mathrm{n}-2},$$

where now y_i stands for the response values of the observation pairs and $\bar{y}$ is the average of such response values.

On another hand, Prediction Intervals (PI's) allow us to establish a band for prediction of future observations, for an specified level of the regressor variable x and a selected confidence level [104]:

$$\hat{\beta}_0 + \hat{\beta}_1 x \pm t_{\alpha/2,n-2}\sqrt{\mathrm{MSR}\left(1 + \frac{1}{\mathrm{n}} + \frac{(\mathrm{x}-\bar{\mathrm{x}})^2}{\sum_{\mathrm{i}=1}^{\mathrm{n}}(\mathrm{x_i}-\bar{\mathrm{x}})^2}\right)}.$$

Clearly, we can observe that the prediction interval at x is always wider than the CI at x, because the prediction interval depends on both the error from the fitted model and the error associated with future observations.

Following [104], it is possible to generalize the previous results for polynomial regression models. In this case, by using simple variable changes, we can analyze the problem by multiple linear regression techniques. We have that a $100(1-\alpha)$ percent CI on the mean response at the point x is:

$$x_0^\top \hat{\beta} \pm t_{\alpha/2,n-p}\sqrt{\mathrm{MSR}\ \mathrm{x}_0^\top(\mathrm{X}^\top\mathrm{X})^{-1}\mathrm{x}_0},$$

where $x_0 \doteq (1, x, x^2, \ldots, x^M)^\top$; $p \doteq M+1$ and $\mathrm{MSR} = \dfrac{\mathrm{y}^\top\mathrm{y} - \hat{\beta}^\top\mathrm{X}^\top\mathrm{y}}{\mathrm{n}-\mathrm{p}}$. On another hand, a $100(1-\alpha)$ percent PI for future observation estimation at the point x is:

$$x_0^\top \hat{\beta} \pm t_{\alpha/2,n-p}\sqrt{\mathrm{MSR}\,(1 + \mathrm{x}_0^\top(\mathrm{X}^\top\mathrm{X})^{-1}\mathrm{x}_0)}.$$

2.4 Prediction Interval-Based Environmental Constrained Economic Dispatch (PIB-ECED)

2.4.1 Prediction-Interval Constrained Programming

As motivation to the approach proposed in this chapter, we begin this section with a simple example. In here just one constraint is considered as uncertain, since the function involved must be experimentally derived. This example allow us to introduce the Prediction-Interval Constrained Programming (PICP), which is the base of the later proposed approach.

Example 2.1. Suppose we are interested in to solve:

$$\begin{aligned} & \min \ f(x_1, x_2) \\ st \qquad & g(x_1, x_2) \leq G_{\max}, \\ & x_1 + x_2 = D, \\ & x_{\min_i} \leq x_i \leq x_{\max_i}, \ i = 1, 2, \end{aligned} \tag{2.4}$$

where the function g must be, somehow, experimentally derived. Particularly, consider the case where $f(x) = 891 + 7.85x_1 + 0.00194x_1^2 + 7.92x_2 + 0.001562x_2^2$; $x_{\min} = (150, 100)^\top$; $x_{\max} = (600, 400)^\top$; $G_{\max} = 250$; $D = 430$ and suppose that we know the true process model:

$$g(x) = 13.860 + 0.3300x_1 + 0.0042x_1^2 + \epsilon, \tag{2.5}$$

where $\epsilon \sim N(0, 10^2)$. Now, using (2.5) and error term specified, we generated random samples of observations. We used these observations to fit the corresponding model, through regression, obtaining results in Table 2.2 (see also Figure 2.1).

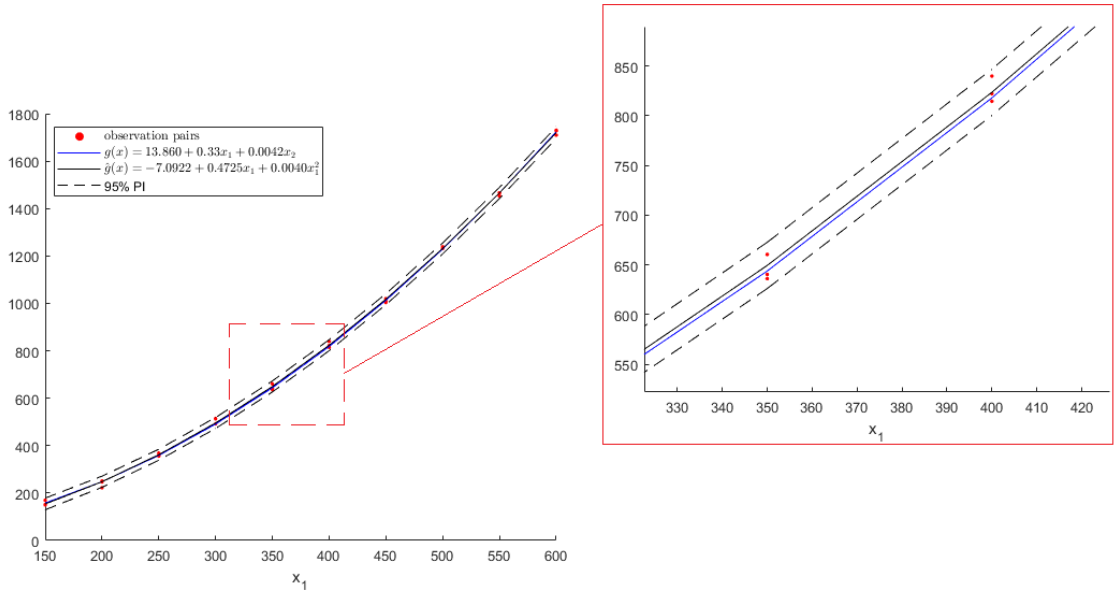


Figure 2.1: Fitted model of Example 1.

Table 2.2: Table of fits Example 1.

Fitted Model	DF	R^2	MSR
$\hat{g}(x) \doteq -7.0922 + 0.4725x_1 + 0.0040x_1^2$	27	0.9996	$(10.9209)^2$

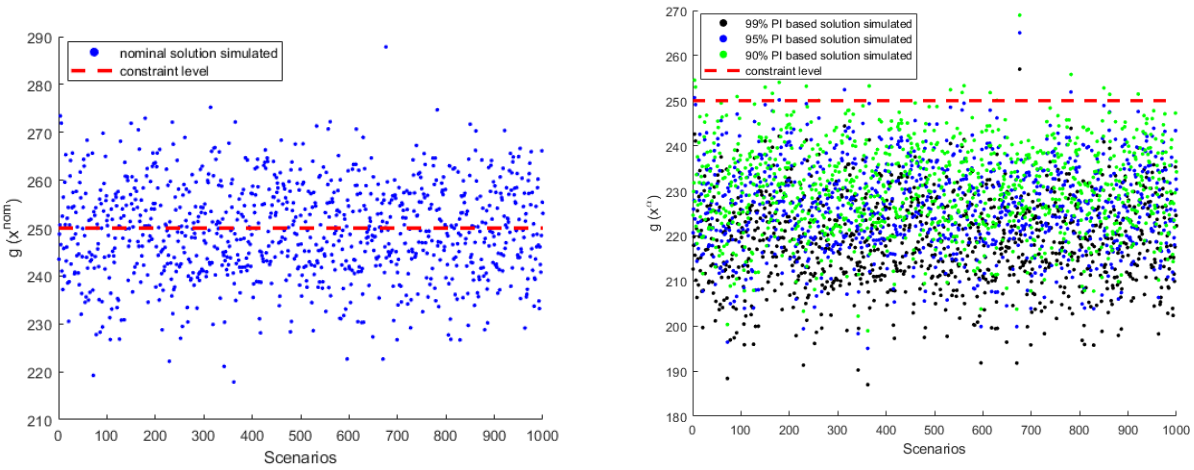
Following the usual deterministic approach, we solve now (2.4) considering the fitted model ($\hat{g}(x)$) without any further consideration of uncertainties associated to parameter estimates.
We obtained the nominal solution:

$$x^{\text{nom}} \doteq (201.06, 228.94)^{\top},$$

which have a cost of $f(x^{\text{nom}}) = 4442.8$. However, although this solution is feasible regarding the mean response, i.e.,

$$E(g(x_{\text{nom}})) = \hat{g}(x_{\text{nom}}) = 249.9907 \leq G_{\max} = 250,$$

it carries a high possibility of become infeasible in some scenarios. We simulated 1000 scenarios considering (2.5), obtaining that violations occurs in 478 times; whereas the mean value of $g(x^{\text{nom}})$ was in fact $249.37 \leq G_{\max}$ over the 1000 scenarios, see Figure 2.2a.



(a) Nominal solution

(b) PI based solutions

Figure 2.2: Solutions performance on the simulated scenarios for Example 1.

Thus, when our main concern is to diminish possibilities of constraint violations on particular situations, and not in average, just solving the nominal problem is not good enough. An alternative to overcome the aforementioned difficulty is provided by the Prediction-Interval Constrained Programming formulation [110, 150]. In such formulation the variability inherent in the regression approach is considered and yields an optimal solution for a specified level of risk. For the example above, the PICP formulation would be:

$$\begin{aligned} & \min \ \ f(x_1, x_2) \\ st \quad & \hat{g}(x_1, x_2) + |t_{\alpha/2, n-3}| \sqrt{\text{MSR}\left(1 + \text{x}_0^{\top} \left(\text{X}^{\top}\text{X}\right)^{-1} \text{x}_0\right)} \leq G_{\max}, \\ & x_1 + x_2 = D, \\ & x_{\min_i} \leq x_i \leq x_{\max_i}, \ i = 1, 2, \end{aligned} \tag{2.6}$$

where the upper limit of the $100(1-\alpha)$ percent PI is additionally considered. We solved (2.6) for several values of α, obtaining the results in Table 2.3.

Table 2.3: PICP solutions for Example 1.

α	x^{α}	Viol. Freq.	Max. Viol. (%)	$f(x^{\alpha})$
0	x^{nom}	478	15.14	4442.8
0.1	$(191.50, 238.50)^{\top}$	20	7.58	4443.2
0.05	$(189.49, 240.51)^{\top}$	5	6.02	4443.3
0.01	$(185.25, 244.75)^{\top}$	1	2.80	4443.8

From Table 2.3 we can see that solving problem (2.6) instead of the nominal (2.4), allow us to obtain solutions with lower risk of become infeasible in some scenario. Clearly, the PICP formulation give us more conservative solutions, but conservativeness can be controled through parameter α, depending on the desicion-maker (DM) risk attitude and experience. Note that, for this particular example, we obtained a considerable diminution in the constraint violations frequency (see Figure 2.2b), with a negligible increase on the objective function value, less than 0.05% in all the cases (for all considered α). Therefore, uncertainties associated to parameter estimation can have great influence on the performance of the optimal solutions; even if the expected (mean) response is adequately represented by the obtained models, as we can appreciate in Figure 2.1 and Table 2.2.

2.4.2 PIB-ECED

In the Environmentally Constrained Economic Dispatch problem, the decision-maker wants to obtain an optimum generation schedule at minimum cost, while polluting emissions are below a pre-specified limit and several other technical constraints are also considered. Then, we must solve:

$$\begin{aligned} & \min \ F(p) \\ st \quad & E(p) \leq E_{\max}, \\ & 1^{\top} p = P_D, \\ & p_{\min} \leq p \leq p_{\max}, \end{aligned} \tag{2.7}$$

where $p = (p_1, p_2, \ldots, p_N)^{\top}$ is the vector of the power output of each generating unit; $F(p) \doteq \sum_{i=1}^{N} f_i(p_i)$ is the total fuel cost; $E(p) \doteq \sum_{i=1}^{N} e_i(p_i)$ is the total emission function; $E_{\max}$ is the maximum allowed emission level; P_D is the power demand and $p_{\min}, p_{\max}$ are the minimum and maximum power output for each generating unit, respectively. To solve this problem the usual approach is to consider the "nominal " ECED formulation (2.7), where all the coefficients of the total fuel cost function and total emission function are considered as fixed (exactly known) values. However, as we already discussed, coefficients obtained through regression are point estimates, therefore, approaches based on statistical intervals could give us better results, since uncertainties in the estimates are then included. Note that in this case, differently to the PICP formulation, uncertainties are involved both in the objective function and in the constraints; but even more important, we have several experimentally derived functions and we are interested in to assess the sum of these functions. To handle uncertainties when sum of experimentally derived functions are involved, several approaches could be used in combination with the PICP formulation. For simplicity in the notation, let us introduce the proposed approaches for a general problem and then

for the ECED problem.

Individual PI's based solution.
In general, consider we are interested in to solve the following problem:

$$\begin{aligned} & \min \ f(x) \\ st \quad & g(x) \leq E_{\max}, \\ & \textstyle\sum_{i=1}^{N} x_i = P_D, \\ & x_{\min_i} \leq x_i \leq x_{\max_i}, \ i = 1, \ldots, N, \end{aligned} \tag{2.8}$$

where functions $f(x) \doteq \sum_{i=1}^{N} f_i(x_i)$ and $g(x) \doteq \sum_{i=1}^{N} g_i(x_i)$ must be experimentally obtained and considered as polynomials. Differently to the usual deterministic approach, that consider coefficients of the estimated models as known (fixed) values, we consider uncertainties on the coefficients estimates through the respectively prediction intervals. Therefore, instead of solving (2.8), perhaps the first and most natural approach would be to solve:

$$\begin{aligned} & \min \ \textstyle\sum_{i=1}^{N} \hat{f}_i(x_i) \\ st \quad & \textstyle\sum_{i=1}^{N} \hat{g}_i(x_i) + |t_{\alpha/2, n_i - M_i - 1}| \sqrt{\mathrm{MSR_i} \left(1 + \mathrm{x_{0_i}} \left(\mathrm{X_i^\top X_i}\right)^{-1} \mathrm{x_{0_i}}\right)} \leq E_{\max}, \\ & \textstyle\sum_{i=1}^{N} x_i = P_D, \\ & x_{\min_i} \leq x_i \leq x_{\max_i}, \ i = 1, \ldots, N, \end{aligned} \tag{2.9}$$

where $\hat{f}_i$, $\hat{g}_i$ are the respectively estimated functions through regression; n_i is the respectively number of observation pairs used for the regression; M_i is the order of the polynomial model fitted; $x_{0_i} = (1, x_i, \ldots, x_i^M)^\top$ and X_i is the corresponding model matrix of the regression. This formulation allow us to obtain solutions immunized, in some sense, against uncertainties associated to models estimates. Note that we are obtaining solutions that remains as feasible, even if the uncertainties are such that the values of $g_i(x_i)$ are in the upper bound of the corresponding $100(1-\alpha)$ percent prediction interval. Clearly, by varying the confidence level, the DM would obtain different solutions: lower values of α give us more conservative solutions, but less likely to be infeasible in some scenario. Note also that in (2.9), we are considering the minimization of the mean value of $f(x)$. However, if uncertainties in the objective function also need to be managed, several formulations could be used motivated by robust optimization theory and DM needs. A possibility could be to minimize the maximum predicted value of function $f(x)$, i.e., to minimize instead the function:

$$\sum_{i=1}^{N} \hat{f}_i(x_i) + |t_{\alpha^f/2, n_i^f - M_i^f - 1}| \sqrt{\mathrm{MSR_i^f} \left(1 + \mathrm{x_{0_i}} \left(\mathrm{X_i^{f\top} X_i^f}\right)^{-1} \mathrm{x_{0_i}}\right)},$$

where we have introduced the f index to identify the regression of each f_i function.

Joint prediction region based solution.
However, since we are interested in the sum of individual responses and not just in each response individually (as in [110, 150, 158]), we need to develop a confidence region for this sum, since

formulation proposed in (2.9) becomes insufficient. Note that we can not effectively assess the risk of constraint violation because, even if we ensure that $P(x_{0_i}^\top \beta_i \in I_i) = 1-\alpha$ for all $i = 1, \ldots, N$; we can expect that $P(x_{0_i}^\top \beta_i \in I_i, \forall i = 1, \ldots, N) < 1-\alpha$. In other words, the $100(1-\alpha)$ percent prediction interval for the sum is not equal to the sum of the individuals $100(1-\alpha)$ prediction intervals. Therefore, as alternatives to ensure an overall required confidence level $(1-\alpha)$, we could use as is usual the Bonferroni's inequality or as proposed in [73], let $[y_{L_i}(x), y_{U_i}(x)]$ be the $(1-\alpha_N)$ prediction interval for $y_i(x)$, then the combination of these individuals PI's:

$$\Omega(x) \doteq [y_{L_1}(x), y_{U_1}(x)] \times \cdots \times [y_{L_N}(x), y_{U_N}(x)],$$

can be regarded as the joint prediction region whose family error rate exactly equals to α. Thus, for $\alpha_N \doteq 1-(1-\alpha)^{1/N}$, we can re-formulate (2.8) considering now the constraint:

$$\sum_{i=1}^{N} \hat{g}_i(x_i) + |t_{\alpha_N/2, n_i - M_i - 1}| \sqrt{\mathrm{MSR_i}\left(1 + \mathrm{x_{0_i}}\left(\mathrm{X_i^\top X_i}\right)^{-1}\mathrm{x_{0_i}}\right)} \leq G_{\max}. \tag{2.10}$$

Let us see a simple, but quite illustrative example where these two approaches are compared.

Example 2.2. Suppose we are interested in to solve the problem:

$$\begin{aligned} & \min \ \ f(x_1, x_2, x_3) \\ st \quad & g(x_1, x_2, x_3) \leq E_{\max}, \\ & x_1 + x_2 + x_3 = P_D, \\ & x_{\min_i} \leq x_i \leq x_{\max_i}, \ i = 1, 2, 3, \end{aligned} \tag{2.11}$$

where $f(x) \doteq x_1^2 + 3x_2 + 10x_3^3$; $g(x_1, x_2, x_3) \doteq g_1(x_1) + g_2(x_2) + g_3(x_3)$; $x_{\min} = (150, 100, 50)^\top$; $x_{\max} = (600, 400, 200)^\top$; $E_{\max} = 3735$; $P_D = 750$ and suppose that we know the true process models:

$$\begin{aligned} g_1(x_1) &\doteq 13.860 + 0.33x_1 + 0.0042x_1^2 + \epsilon_1, \\ g_2(x_2) &\doteq 561.0 + 7.92x_2 + 0.001562x_2^2 + \epsilon_2, \\ g_3(x_3) &\doteq 78.0 + 7.97x_3 + 0.00482x_3^2 + \epsilon_3, \end{aligned} \tag{2.12}$$

with $\epsilon_1 \sim N(0, 10^2)$; $\epsilon_2 \sim N(0, 14^2)$ and $\epsilon_3 \sim N(0, 6^2)$. We generated random observation pairs using (2.12), obtaining results in Table (2.4).

Table 2.4: Table of fits Example 2.

Fitted Model	DF	R^2	MSR
$\hat{g}_1(x) \doteq -7.0922 + 0.4725x_1 + 0.0040x_1^2$	27	0.9995	$(10.9209)^2$
$\hat{g}_2(x) \doteq 531.6668 + 8.2192x_2 + 0.0010x_2^2$	27	0.9997	$(15.2892)^2$
$\hat{g}_3(x) \doteq 65.4287 + 8.2265x_3 + 0.0038x_3^2$	27	0.9998	$(6.5525)^2$

In the usual deterministic approach, we solve (2.11) considering the estimates models from Table (2.4), obtaining the nominal solution:

$$x^{\text{nom}} = (551.04, 148.96, 50.00)^\top,$$

which have a cost $f(x^{\text{nom}}) = 1554087.17$. As discussed previously, this solution, although remains as feasible considering the expected response, carries a high possibility of become infeasible in some scenarios, since uncertainties on the estimates coefficients have not been considered. In Figure 2.3 we can observe the performance of the nominal solution when 1000 scenarios were simulated using the true models (2.12). In this case violation occurs 517 times.

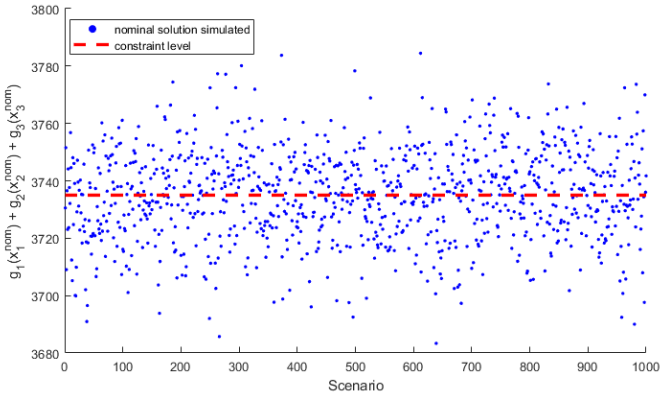


Figure 2.3: Nominal solution performance on the simulated scenarios for Example 2.

Now, in order to handle uncertainties and provide more robust solutions, we re-formulate (2.11) considering each individual 95% ($\alpha = 0.05$) prediction interval, then we need to solve:

$$\begin{aligned} &\min \ \ f(x_1, x_2, x_3) \\ st \quad &\textstyle\sum_{i=1}^{3} \hat{g}_i(x) + |t_{0.025, n_i - 3}| \sqrt{\text{MSR}_\text{i} \left(1 + \text{x}_{0_\text{i}} \left(\text{X}_\text{i}^\top \text{X}_\text{i}\right)^{-1} \text{x}_{0_\text{i}}\right)} \leq E_{\max}, \\ &\textstyle\sum_{i=1}^{3} x_i = P_D, \\ &x_{\min_\text{i}} \leq x_i \leq x_{\max_\text{i}}, \ \ i = 1, 2, 3. \end{aligned} \tag{2.13}$$

We obtained as solution:

$$x^{\alpha} = (571.23, 128.77, 50.0000)^\top,$$

which have a cost $f(x^{\alpha}) = 1576690.39$. In Figure 2.4 we can observe the performance of the solution obtained in the 1000 scenarios simulated. Note that no violations of the constraint occur in this case, whereas the increment in the cost was negligible, around 1.5%.

As another alternative of solution, consider now the joint prediction region approach. In this case, for an overall $\alpha = 0.05$ we must enhance each individual prediction interval considering $\alpha_N \doteq 1 - (1 - \alpha)^{1/N} = 0.01695$. Solving (2.13) for α_N we obtain as solution:

$$x^{\alpha_N} = (576.42, 123.58, 50.00)^\top,$$

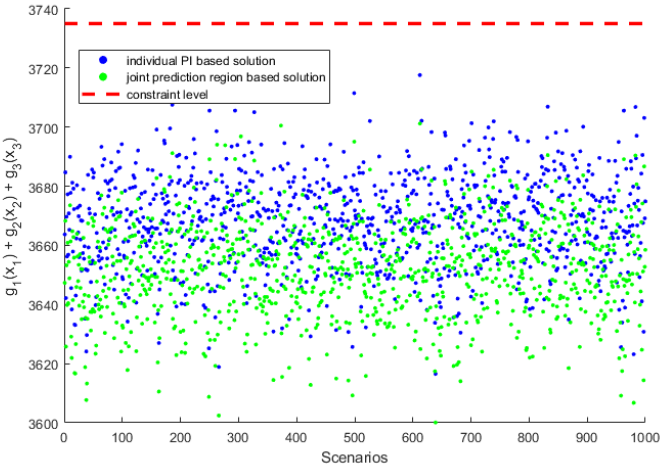


Figure 2.4: Prediction interval-based solutions performance on the simulated scenarios Example 2.

which have a cost $f(x^{\alpha_N}) = 1582626.83$. This represent an increase of 1.84%, regarding the nominal cost; but as we can see in Figure 2.4 no violation occurs when x^{α_N} was implemented as solution. Then, for this problem, we diminished the risk of constraint violations by considering the uncertainties associated to the experimental derivation of the involved functions. Note that we obtained solutions that remains as feasible in most scenarios and for this particular example, the increase of the optimal value obtained was negligible in all the cases.
However, the main limitation with the approach proposed in (2.10) is that, when N increases, $\alpha_N \to 1$, implying the need of extremely large individual PI's, making this approach useless (same difficulty faced if we use Bonferroni inequality). In such cases, optimal solutions, if there exist, need to be very conservatives, since the terms $|t_{\alpha_N/2, n_i - M_i - 1}|$ becomes bigger. As alternative, for moderates values of N, the DM could diminish the overall confidence level required $(1-\alpha)$ and analyze if solutions obtained still perform well under uncertainties; but this strategy remains useless for great values of N. On another hand, the DM can try to use formulation provided in (2.9) and analyze the solutions obtained. Note that although this formulation have not a clear relationship between individual confidence level and the overall confidence level required, still uncertainties on the estimation of the models are considered. Thus, this approach can provide solutions with better performance than the nominal ones, diminishing risk of constraint violations.

Approximated PI based solution.
To overcome the aforementioned disadvantages, we propose another alternative formulation to deal with uncertainties of several experimentally derived functions. This approach is mainly based on the results on [151, 152], about significance of the difference between means when several different population variances are involved; and particularly, in the application to derive confidence statistics over the sum of component regressions, as made in [145]. For the sake of completeness we present here some of the relevant results of [151, 152].
Let η be a population parameter which is estimated by an observed quantity y, normally distributed with variance σ_y^2. Let $\sigma_y^2 = \sum_{i=1}^{N} \lambda_i \sigma_i^2$ where the λ_i, are known positive numbers and the σ_i^2 are unknown variances. Suppose that the observed data provide estimates $\hat{\sigma}_i^2$ of these variances, based

on f_i degrees of freedom, respectively. In such case, an approximation to the distribution of $\sum_{i=1}^{N} \lambda_i \hat{\sigma}_i^2$, using a Type III curve with start, mean and variance suitably adjusted, becomes useful. We consider $\sum_{i=1}^{N} \lambda_i \hat{\sigma}_i^2$ of the form $\sum_{i=1}^{N} a_i \chi_i^2$, where $a_i = \frac{\lambda_i \sigma_i^2}{f_i}$ and each χ_i^2 is independently distributed as a χ^2 variate. For simplicity the case $N = 2$ is considered, but extension for $N \in \mathbb{N}$ is straightforward. Let $w \doteq a_1\chi_1^2 + a_2\chi_2^2$, and the approximate distribution curve written in the form:

$$p(w)dw = \frac{1}{\Gamma\left(\frac{1}{2}f\right)} e^{-\frac{w}{2g}} \left(\frac{w}{2g}\right)^{\frac{f}{2}-1} d\left(\frac{w}{2g}\right). \tag{2.14}$$

Then, making the first two moments of (2.14):

$$\begin{aligned} \text{mean} &= fg, \\ \text{variance} &= 2g^2 f, \end{aligned}$$

agree with the two moments of w:

$$\begin{aligned} \text{mean} &= a_1 f_1 + a_2 f_2, \\ \text{variance} &= 2(a_1^2 f_1 + a_2^2 f_2), \end{aligned}$$

we obtain that $f = \frac{(a_1 f_1 + a_2 f_2)^2}{a_1^2 f_1 + a_2^2 f_2}$ and $\mathrm{g} = \frac{\mathrm{a_1^2 f_1 + a_2^2 f_2}}{\mathrm{a_1 f_1 + a_2 f_2}}$. Therefore, w/g is approximately distributed as χ^2 with degree of freedom f. This particularly allow us to say that the criterion:

$$v \doteq \frac{y - \eta}{\sqrt{\lambda_1 \hat{\sigma}_1^2 + \lambda_2 \hat{\sigma}_2^2}}$$

follows approximately the Student's t distribution with degree of freedom:

$$f = \frac{(\lambda_1 \sigma_1^2 + \lambda_2 \sigma_2^2)^2}{\frac{\lambda_1^2 \sigma_1^4}{f_1} + \frac{\lambda_2^2 \sigma_2^4}{f_2}}.$$

More generally, for $N \in \mathbb{N}$, the same line of arguments leads to say that the criterion:

$$v = \frac{y - \eta}{\sqrt{\sum_{i=1}^{N} \lambda_i \hat{\sigma}_i^2}}$$

is approximately distributed as Student's t with degree of freedom:

$$f = \frac{(\sum_{i=1}^{N} \lambda_i \sigma_i^2)^2}{\sum_{i=1}^{N} \lambda_i^2 \sigma_i^4 / f_i}. \tag{2.15}$$

Not knowing the population variances σ_i's in (2.15), Welch [152] proposed to use:

$$f = \frac{\left(\sum_{i=1}^{N} \lambda_i \hat{\sigma}_i^2\right)^2 - 2\left(\sum_i^N \lambda_i^2 \hat{\sigma}_i^4/(f_i + 2)\right)}{\sum_{i=1}^{N} \lambda_i^2 \hat{\sigma}_i^4/(f_i + 2)}, \tag{2.16}$$

since it may be shown that the numerator of (2.16) has, in repeated samples, an average value $(\sum_{i=1}^{N} \lambda_i \sigma_i^2)^2$, and the denominator has average value $\sum_{i=1}^{N} \lambda_i^2 \sigma_i^4 / f_i$. In a certain sense, therefore, (2.16) is a fair estimate of (2.15).

In order to apply these results to the general problem (2.8), note that we want to estimate, for N individual given predictors $\bar{x} \doteq (\bar{x}_1, \ldots, \bar{x}_N)^\top$, $g(\bar{x}) \doteq \sum_{i=1}^N g_i(\bar{x}_i)$, where each $g_i = X_i\beta_i + e_i$, $i = 1, \ldots, N$; by using the sum of the fitted values given by $\hat{g}(\bar{x}) \doteq \sum_{i=1}^N \hat{g}_i(\bar{x}_i)$, with $\hat{g}_i = X_i\hat{\beta}_i$. Therefore we know, from model's assumptions and multiple linear regression theory, that the variable:

$$u \doteq \frac{g(\bar{x}) - \hat{g}(\bar{x})}{\sqrt{\sum_{i=1}^N \sigma_i^2 \left(1 + \bar{x}_{0_i}^\top \left(X_i^\top X_i\right)^{-1} \bar{x}_{0_i}\right)}} = \frac{g(\bar{x}) - \hat{g}(\bar{x})}{\sqrt{\sum_{i=1}^N \sigma_i^2 \lambda_i}} \sim N(0,1),$$

where $\bar{x}_{0_i} = (1, \bar{x}_i, \bar{x}_i^2, \ldots, \bar{x}_i^M)^\top$ and $\lambda_i \doteq 1 + \bar{x}_{0_i}^\top \left(X_i^\top X_i\right)^{-1} \bar{x}_{0_i}$. Defining as before $w \doteq \sum_{i=1}^N \lambda_i \hat{\sigma}_i^2 = \sum_{i=1}^N a_i \chi_i^2$; we have that $\frac{\sqrt{fg}u}{\sqrt{w}} \sim t_f$, where now:

$$f = \frac{\left(\sum_{i=1}^N a_i f_i\right)^2}{\sum_{i=1}^N a_i^2 f_i},\ g = \frac{\sum_{i=1}^N a_i^2 f_i}{\sum_{i=1}^N a_i f_i}, a_i = \frac{\lambda_i \sigma_i^2}{f_i} \text{ and } \mathrm{f_i} = \mathrm{n_i} - \mathrm{M_i} - 1.$$

Therefore,

$$\frac{\sqrt{fg}u}{\sqrt{w}} = \frac{(g(\bar{x}) - \hat{g}(\bar{x}))\sqrt{\sum_{i=1}^N a_i f_i}}{\sqrt{\sum_{i=1}^N \sigma_i^2 \lambda_i}\sqrt{\sum_{i=1}^N \hat{\sigma}_i^2 \lambda_i}} = \frac{g(\bar{x}) - \hat{g}(\bar{x})}{\sqrt{\sum_{i=1}^N \mathrm{MSR_i} \lambda_\mathrm{i}}} \sim t_f, \tag{2.17}$$

where the effective degree of freedom is obtained from:

$$f = \frac{\left(\sum_{i=1}^N \lambda_i \sigma_i^2\right)^2}{\sum_{i=1}^N \lambda_i^2 \sigma_i^4 / f_i} = \frac{\left(\sum_{i=1}^N \sigma_i^2 \left(1 + \bar{x}_{0_i}^\top \left(X_i{}^\top X_i\right)^{-1} \bar{x}_{0_i}\right)\right)^2}{\sum_{i=1}^N \sigma_i^4 \left(1 + \bar{x}_{0_i}^\top \left(X_i{}^\top X_i\right)^{-1} \bar{x}_{0_i}\right)^2 / (n_i - M_i - 1)}. \tag{2.18}$$

As before, an approximation to (2.18) can be obtained using:

$$\begin{aligned} f &= \frac{\left(\sum_{i=1}^N \hat{\sigma}_i^2 \left(1 + \bar{x}_{0_i}^\top \left(X_i{}^\top X_i\right)^{-1} \bar{x}_{0_i}\right)\right)^2 - 2\left(\sum_{i=1}^N \hat{\sigma}_i^4 \left(1 + \bar{x}_{0_i}^\top \left(X_i{}^\top X_i\right)^{-1} \bar{x}_{0_i}\right)^2 / (n_i - M_i + 1)\right)}{\sum_{i=1}^N \hat{\sigma}_i^4 \left(1 + \bar{x}_{0_i}^\top \left(X_i{}^\top X_i\right)^{-1} \bar{x}_{0_i}\right)^2 / (n_i - M_i + 1)} \\ &= \frac{\left(\sum_{i=1}^N \mathrm{MSR_i} \left(1 + \bar{\mathrm{x}}_{0_\mathrm{i}}^\top \left(\mathrm{X_i}{}^\top \mathrm{X_i}\right)^{-1} \bar{\mathrm{x}}_{0_\mathrm{i}}\right)\right)^2 - 2\left(\sum_{i=1}^N \mathrm{MSR_i}{}^2 \left(1 + \bar{x}_{0_i}^\top \left(X_i{}^\top X_i\right)^{-1} \bar{x}_{0_i}\right)^2 / (n_i - M_i + 1)\right)}{\sum_{i=1}^N \mathrm{MSR_i}{}^2 \left(1 + \bar{x}_{0_i}^\top \left(X_i{}^\top X_i\right)^{-1} \bar{x}_{0_i}\right)^2 / (n_i - M_i + 1)}. \end{aligned} \tag{2.19}$$

Therefore, a prediction interval for $g(x) = \sum_{i=1}^N g_i(x_i)$, given N predictors $x \doteq (x_1, \ldots, x_N)^\top$, can be obtained from (2.17), (2.19):

$$\sum_{i=1}^N \hat{g}_i(x_i) \pm t_{\alpha/2,f} \sqrt{\sum_{k=1}^N \mathrm{MSR_k} \left(1 + \mathrm{x}_{0_\mathrm{k}}^\top \left(\mathrm{X_k}{}^\top \mathrm{X_k}\right)^{-1} \mathrm{x}_{0_\mathrm{k}}\right)}.$$

We point out that a simple approximation to the effective degree of freedom f can be obtained as in [39, 145]. In this case to approximate (2.15), instead of (2.16), we use that:

$$f = \frac{(\sum_{i=1}^{N} \lambda_i \hat{\sigma}_i^2)^2}{\sum_{i=1}^{N} \lambda_i^2 \hat{\sigma}_i^4 / f_i} = \frac{(\sum_{i=1}^{N} \lambda_i \mathrm{MSR_i})^2}{\sum_{i=1}^{N} \lambda_i^2 \mathrm{MSR_i}^2 / (n_i - M_i - 1)} \doteq \bar{f}. \tag{2.20}$$

Thus, we obtain the approximated prediction interval:

$$\sum_{i=1}^{N} \hat{g}_i(x_i) \pm t_{\alpha/2,\bar{f}} \sqrt{\sum_{k=1}^{N} \mathrm{MSR_k} \left(1 + \mathrm{x}_{0_\mathrm{k}}^{\top} \left(\mathrm{X_k}^{\top} \mathrm{X_k}\right)^{-1} \mathrm{x}_{0_\mathrm{k}}\right)}. \tag{2.21}$$

Thus, for the general problem (2.8), when N increases, we could use the proposed approximated prediction intervals and solve then:

$$\begin{gathered} \min \ \hat{f}(x) \\ st \quad \sum_{i=1}^{N} \hat{g}_i(x_i) + |t_{\alpha/2,\bar{f}(x)}| \sqrt{\sum_{k=1}^{N} \mathrm{MSR_k} \left(1 + \mathrm{x}_{0_\mathrm{k}}^{\top} \left(\mathrm{X_k}^{\top} \mathrm{X_k}\right)^{-1} \mathrm{x}_{0_\mathrm{k}}\right)} \leq E_{\max} \\ \sum_{i=1}^{N} x_i = P_D, \\ x_{\min_\mathrm{i}} \leq x_i \leq x_{\max_\mathrm{i}}, \ i = 1, \ldots, N. \end{gathered} \tag{2.22}$$

Note that, whatever we choose to use (2.19) or (2.20) as the effective degree of freedom in (2.22), we just have one percentile of the Student's t distribution involved ($t_{\alpha/2,f(x)}$) and not N as in (2.9) or (2.10). However, this single value relates the degree of freedom and the mean squared error on the regression of each function g_i; as well as the specific value of the predictor where the prediction interval need to be estimated. Then, in order to solve (2.22) we need to handle the implicit dependency between the variable to optimize (x) and the Student's t inverse cumulative distribution function. This makes problem (2.22) more complicated than problems (2.9) or (2.10), from the computational point of view. Optimization algorithms and strategies to efficiently solve (2.22) will be analyzed in a forthcoming paper, but alternatives could be, for example, to use existing quantile approximations for the Student's t distribution, see [53, 84, 111, 127]. Nonetheless, for the Example 2 and the 4-generating units system analyzed in Section 2.5, just by using MATLAB's incorporated routines, we were able to obtain solutions to the proposed formulation (2.22), by considering the approximation (2.20).

Particularly for the Example 2, by using formulation proposed in (2.22), we obtain as solution:

$$x_{\bar{f}} = (562.65, 137.35, 50.00)^{\top},$$

which have a cost $f(x_{\bar{f}}) = 1566990.69$. As we can see in Figure 2.5, this solution have a low risk of constraint violations, whereas have a cost less conservative than solutions x^{α} and x^{α_N}. In this case the cost increase around 0.83%, compared with the nominal cost and violations just occurs 18 times in the 1000 scenarios.

Therefore, by solving (2.22), we could effectively obtain robust solutions with low risk of constraint violations, while conservativeness of the solutions can be now effectively managed through parameter α.

Then, considering all the exposed above, as alternative to handle uncertainties on the Environmental Constrained Economic Dispatch problem, the decision-maker can solve the following problem:

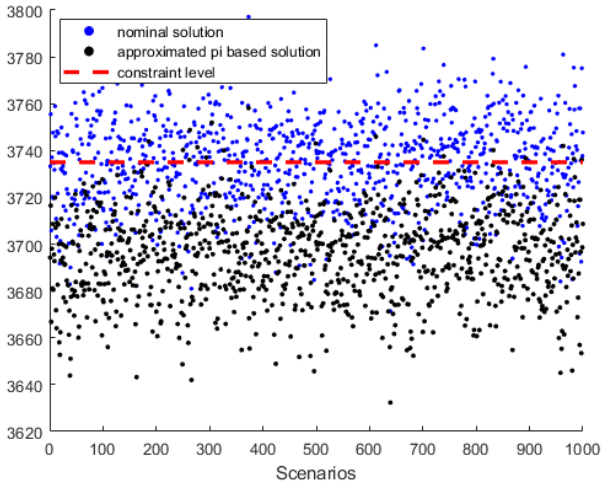


Figure 2.5: Solutions performance on the simulated scenarios for Example 2.

$$\begin{aligned} &\min \ F_{\alpha f}(p) \\ st \ \ &E_{\alpha f}(p) \leq E_{\max} \\ &1^\top p = P_D, \\ &p_{\min} \leq p \leq p_{\max}, \end{aligned} \tag{2.23}$$

where:

$$F_{\alpha_f}(p) \doteq \sum_{i=1}^{N}\sum_{j=1}^{M} \hat{a}_{0i} + \hat{a}_{ji} p_i^j + |t_{\alpha/2, f_F(p)}| \sqrt{\sum_{k=1}^{N} \mathrm{MSR}_{\mathrm{F}}^{\mathrm{k}} \left(1 + \mathrm{p}_{0_{\mathrm{k}}}^\top \left(\mathrm{X}_{\mathrm{F}}^{\mathrm{k}\top} \mathrm{X}_{\mathrm{F}}^{\mathrm{k}}\right)^{-1} \mathrm{p}_{0_{\mathrm{k}}}\right)},$$

$$E_{\alpha f}(p) \doteq \sum_{i=1}^{N}\sum_{j=1}^{M} \hat{\alpha}_{0i} + \hat{\alpha}_{ji} p_i^j + |t_{\alpha/2, f_E(p)}| \sqrt{\sum_{k=1}^{N} \mathrm{MSR}_{\mathrm{E}}^{\mathrm{k}} \left(1 + \mathrm{p}_{0_{\mathrm{k}}}^\top \left(\mathrm{X}_{\mathrm{E}}^{\mathrm{k}\top} \mathrm{X}_{\mathrm{E}}^{\mathrm{k}}\right)^{-1} \mathrm{p}_{0_{\mathrm{k}}}\right)},$$

where we have used $f_F(p)$ and $f_E(p)$ to identify the effective degree of freedom for the total fuel cost and total emission generated, respectively.
Note that in (2.23) a *minmax* formulation have been used for the objective function, i.e., we are minimizing the maximum predicted fuel cost for a given confidence level. However, if uncertainties are not to be considered in the objective function, the DM can simple minimize the expected (mean) total fuel cost instead:

$$F(p) = \sum_{i=1}^{N}\sum_{j=1}^{M} \hat{a}_{0i} + \hat{a}_{ji} p_i^j. \tag{2.24}$$

The results presented in this section are particularly useful to deal with worst-case (pessimistic) scenario approach. Consider polluted areas, where emission limits for power plants are enforced to ensure air quality, air quality that is influenced by changing meteorological conditions [62, 63].

Thus, for particular planning periods, we are interested in to find an optimal solution to the ECED problem that remains environmentally feasible, or at least with low risk of becoming unfeasible. In this case, a stochastic approach (as those discussed in Section 2.2) is not fully satisfactory. Note that in such investigations optimal solutions are obtained regarding mean responses, therefore, as pointed out in [110], solutions will be optimal or even feasible on average, but will carry a large degree of risk of not meeting environmental limits when the obtained schedule is implemented. This phenomenon becomes evident in the next section, where we analyzed a system consisting on 4-generating units from the Chilean electrical power network. We compared and analyzed the deterministic and robust solutions obtained.

2.5 Study-cases: 4-generating units system from the Chilean electrical power network

In this section we analyze the effect of coefficients uncertainties in a real system. We use operating data of four coal-fired generating units from the Chilean electric power network for, by firstly estimating fuel cost and emission functions through regression, solving the nominal ECED problem. Then, we consider the respectively prediction intervals to obtain robust solutions with lower risk of become infeasible in some scenario. We particularly analyzed SO_2 emissions and fuel consumption. In order to obtain the fuel cost, consumption must be multiplied by the fuel's price. We were not able to obtain this value, but the analysis is similar, since the optimal solutions coincide if we consider fuel consumption or fuel cost.

Table 2.5: Table of fit fuel consumption functions.

i	Unit	$\hat{a}_{0_i}$	$\hat{a}_{1_i}$	DF	RMSR	R^2
1	Bocamina U1	8.0204	0.3575	6	2.5825	0.9720
2	Bocamina U2	1.0490	0.3634	8	3.7902	0.9908
3	Campiche	2.0698	0.4010	9	3.0707	0.9919
4	Santa María	10.7697	0.3393	6	10.7	0.94

Table 2.6: Table of fit SO_2 emission functions.

i	Unit	$\hat{\alpha}_{0_i}$	$\hat{\alpha}_{1_i}$	$\hat{\alpha}_{2_i}$	$\hat{\alpha}_{3_i}$	DF	RMSR	R^2
1	Bocamina U1	174.1	-5.77	0.3023	-0.001369	4	102.6	0.97
2	Bocamina U2	-10.93	-4.832	0.09296	-0.0001386	6	230.6	0.98
3	Campiche	-2.941	28.11	-0.1505	0.0003207	7	249.5	0.93
4	Santa María	41.92	-3.363	0.03252	0	5	350.8	0.92

In Tables 2.5 and 2.6 we can observe the parameters obtained when multiple linear regression was used. Note that a linear model ($f_i(p_i) = \hat{a}_{0_i} + \hat{a}_{1_i} p_i$) represents well relationship between power output and fuel consumption for each generating unit; whereas quadratic ($e_i(p_i) = \hat{\alpha}_{0_i} + \hat{\alpha}_{1_i} p_i + \hat{\alpha}_{2_i} p_i^2$) and cubic models ($e_i(p_i) = \hat{\alpha}_{0_i} + \hat{\alpha}_{1_i} p_i + \hat{\alpha}_{2_i} p_i^2 + \hat{\alpha}_{3_i} p_i^3$), seems to correctly represent relationship

between power output and SO_2 emissions. In Figure 2.6 we can see the fitted models (using the observation pairs available) and its corresponding 95% prediction intervals.
Now, considering the above models' parameter estimates, we solved the ECED problem and several scenarios were simulated depending on different power demands and emission limits. We compared the results when a deterministic approach was used (considering coefficients estimates as exacts) against results obtained when uncertainties on coefficients estimated are considered. The scenarios were created considering variance in the response given by the respectively mean squared error computed. All the results we present in here were obtained by using MATLAB built-in functions on a personal computer.

2.5.1 Case 1: $P_D = 600$ MW and $E_{max} = 10940$ kg

For this case consider that the power demand is $P_D = 600$ MW; maximum emission allowed during the period $E_{\max} = 10940$ kg; $p_{\min} = (70, 240, 140, 50)^\top$ and $p_{\max} = (128, 350, 270, 350)^\top$. If not uncertainties are considered on the coefficients estimates (most usual approach), then we need to solve the nominal problem (2.7), that for our case becomes:

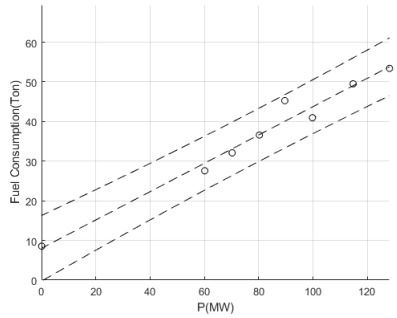
$$\begin{aligned} &\min \ \textstyle\sum_{i=1}^{4} \hat{a}_{0i} + \hat{a}_{1i} p_i \\ st \quad &\textstyle\sum_{i=1}^{4} \hat{\alpha}_{0_i} + \hat{\alpha}_{1_i} p_i + \hat{\alpha}_{2_i} p_i^2 + \hat{\alpha}_{3_i} p_i^3 \leq E_{\max}, \\ &\textstyle\sum_{i=1}^{4} p_i = P_D, \\ &p_{\min_i} \leq p_i \leq p_{\max_i}, \end{aligned} \tag{2.25}$$

where coefficients are in Table 2.5 and Table 2.6. In this case, when we solve (2.25), we obtained as optimal nominal solution $p^{\text{nom}} = (70, 240, 140, 150)^\top$. This solution satisfy all the constraints in (2.25).
Now, we analyze if considering uncertainties associated to emission functions coefficients estimates, have any effect on the optimal solution. For simplicity and to provide fair comparisons, we considered the same objective function and try to solve now the problems:

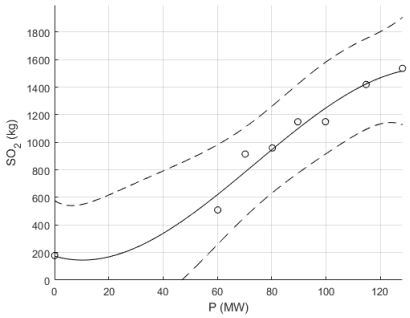
$$\begin{aligned} &\min \ \textstyle\sum_{i=1}^{4} \hat{a}_{0i} + \hat{a}_{1i} p_i \\ st \quad &\textstyle\sum_{i=1}^{4} \hat{e}_i(p_i) + |t_{\alpha/2, n_i - M_i - 1}| \sqrt{\mathrm{MSR_i} \left(1 + \mathrm{p_{0_i}} \left(\mathrm{X_i^\top X_i}\right)^{-1} \mathrm{p_{0_i}}\right)} \leq E_{\max}, \\ &\textstyle\sum_{i=1}^{4} p_i = P_D, \\ &p_{\min_i} \leq p_i \leq p_{\max_i}, \end{aligned} \tag{2.26}$$

and

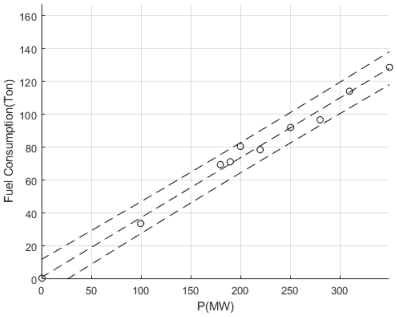
$$\begin{aligned} &\min \ \textstyle\sum_{i=1}^{4} \hat{a}_{0i} + \hat{a}_{1i} p_i \\ st \quad &\textstyle\sum_{i=1}^{4} \hat{e}_i(p_i) + |t_{\alpha/2, \bar{f}(p)}| \sqrt{\sum_{k=1}^{4} \mathrm{MSR_k} \left(1 + \mathrm{p_{0_k}} \left(\mathrm{X_k^\top X_k}\right)^{-1} \mathrm{p_{0_k}}\right)} \leq E_{\max}, \\ &\textstyle\sum_{i=1}^{4} p_i = P_D, \\ &p_{\min_i} \leq p_i \leq p_{\max_i}, \end{aligned} \tag{2.27}$$



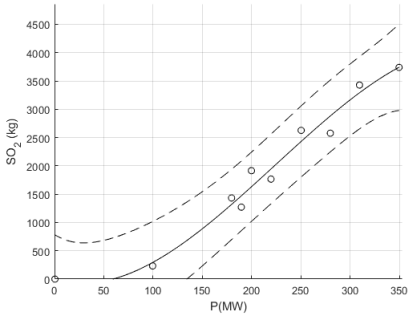
(a) Bocamina U1 fuel cost



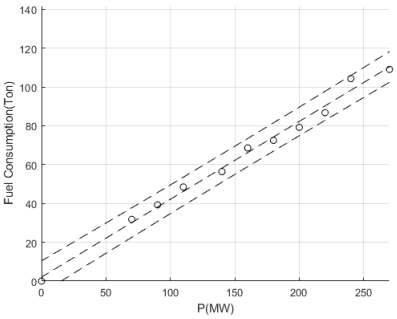
(b) Bocamina U1 SO_2 emissions



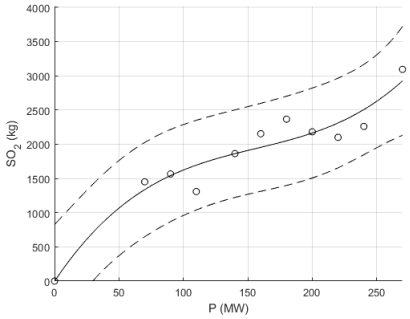
(c) Bocamina U2 fuel cost



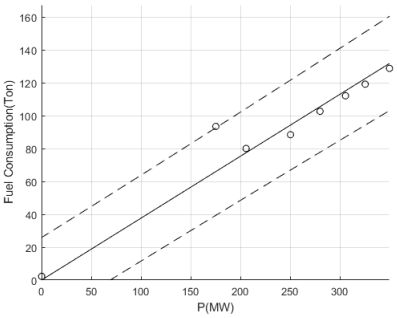
(d) Bocamina U2 SO_2 emissions



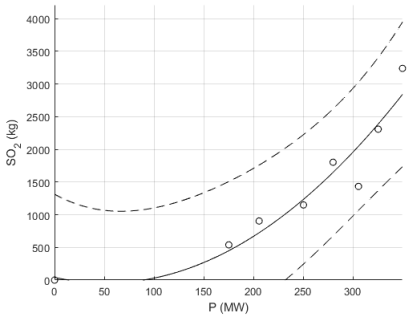
(e) Campiche fuel cost



(f) Campiche SO_2 emissions



(g) Santa María fuel cost



(h) Santa María SO_2 emissions

Figure 2.6: Fuel cost and SO_2 emissions regression.

where $\hat{e}_i$ is the fitted emission function of each generating unit. We will denoted solutions of (2.26) as p^{α}, whereas solutions of (2.27) will be denoted as $p_{\bar{f}}$. In this case we obtain the same optimal solution ($p^{\alpha} = p^{\text{nom}}$) either if we consider the 95% percent individual or overall prediction intervals; i.e., solving (2.26) considering $\alpha = 0.05$ or $\alpha_N = 0.0127$ (as proposed in (2.10)). On another hand, for this small power system we were able to efficiently solve formulation proposed in (2.27), just by using the built-in routines in MATLAB. The same solution was obtained ($p_{\bar{f}} = p^{\text{nom}}$). Therefore, uncertainties have not effect on the ECED problem in this situation. As we can see in Figure 2.7, the maximum emission level allowed is considerable bigger than the predicted (mean) total emissions simulated, even if uncertainties on model estimates are considered. Therefore, violations are not probably to occur and the deterministic formulation provides suitable solutions.

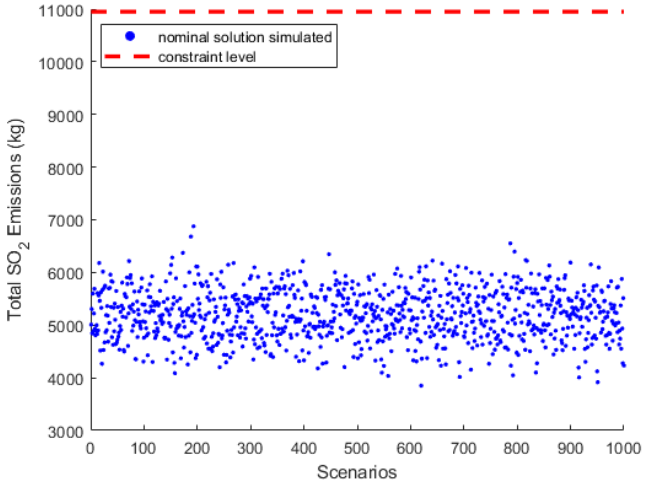


Figure 2.7: SO_2 emission scenarios simulated ($P_D = 600$ MW/$E_{max} = 10940$ kg).

2.5.2 Case 2: $P_D = 600$ MW and $E_{max} = 8940$ kg

For this case, if the power demand $P_D = 600$ MW is maintained, but now the maximum emission level is diminished a little, we obtained the same nominal solution $p^{\text{nom}} = (70, 240, 140, 150)^{\top}$) that in the previous case. Even more, we obtain the same solutions if formulation provided in (2.26) and (2.27) are used considering $\alpha = 0.05$. However, the formulation provided in (2.10) that considers the joint prediction region, becomes too conservative (as previously discussed) in this case, even considering that a small number of generating units $N = 4$ are to be dispatched. In this case, no feasible solution was obtained when we attempt to solve (2.26) for $\alpha_N = 0.0127$.

2.5.3 Case 3: $P_D = 850$ MW and $E_{max} = 8940$ kg

Let us suppose now that the maximum SO_2 emission level is the same $E_{max} = 8940$ kg, but the power demand is increased $P_D = 850$ MW.
Solving (2.25), we obtained the nominal solution $p^{\text{nom}} = (120, 240, 140, 350)^{\top}$, which have an expected total fuel cost $F(p^{\text{nom}}) = 326.9185$. In this case no feasible solution exist considering $\alpha = 0.05$ both for the individual or joint prediction interval-based formulations (2.9) and (2.10).

Table 2.7: Comparison between deterministic and PI based robust solutions($P_D = 850$ MW/$E_{max} = 8940$ kg).

$\alpha = 0.05$	p_1	p_2	p_3	p_4	$F(p)$	$F(p)/F(p^{\text{nom}})$	$E(p)$	Viol. Freq.	Max. Viol.
p^{nom}	120.00	240.00	140.00	350.00	326.9185	1	8448.19	168	9848.07
p^{α}	-	-	-	-	-	-	-	-	
p^{α_N}	-	-	-	-	-	-	-	-	
$p_{\bar{f}}$	70.98	240.00	210.97	328.05	330.4026	1.0107	7743.90	7	9143.78

However, by using the proposed formulation based on the approximated prediction interval (2.27), we obtain as solution $p_{\bar{f}} = (70.98, 240.00, 210.97, 328.05)^\top$, which have an expected total fuel cost $F(p_{\bar{f}}) = 330.4026$. This represent an increase of 1.07% respect to the cost of the nominal solution, but as we can see in Figure 2.8 and Table 2.7, the risk of environmental violation is diminished. In the 1000 simulated scenarios, violations occur 168 times (with a maximum violation of 10.15%) when the nominal solution was selected; whereas just 7 violations occur when $p_{\bar{f}}$ was selected (with a maximum violation of 2.28%). Therefore, with a negligible increase in the total fuel cost, we managed to significantly diminish (from 16.8% to 0.7%) the risk of environmental violations in this situation.

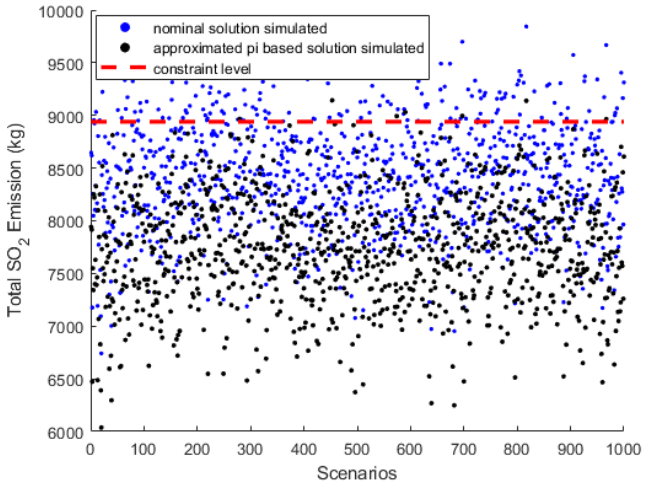
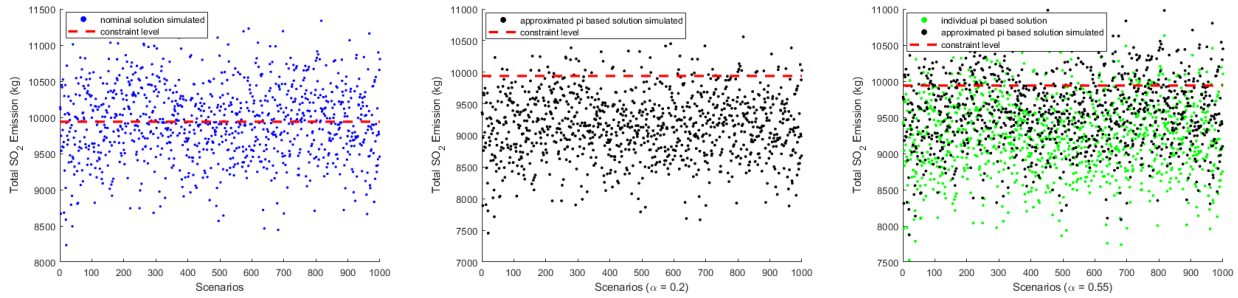


Figure 2.8: SO_2 emission scenarios simulated ($P_D = 850$ MW/$E_{max} = 8940$ kg).

2.5.4 Case 4: $P_D = 980$ MW and $E_{max} = 9940$ kg

Suppose that $P_D = 980$ MW and $E_{max} = 9940$ kg. Solving the nominal problem we obtain the solution $p^{\text{nom}} = (128.00, 350.00, 158.03, 343.97)^\top$, which have a expected total fuel cost $F(p^{\text{nom}}) = 374.9374$. In this case for $\alpha = 0.05$ no feasible solution exist for any of the formulations (2.26), (2.27). However, by varying the confidence level the DM could still obtain solutions with better robustness than the nominal one.
In Table 2.8 we can observe that in general with the joint region approach, feasible solutions can



(a) Nominal solution (b) $\alpha = 0.2$ (c) $\alpha = 0.55$

Figure 2.9: Simulated scenarios for several confidence levels.

not be obtained. This occurs because the proposed formulation becomes extremely conservative, even considering a small number of generating units. In the case of the individual prediction interval approach, for lower values of α ($\alpha = 0.1$ and $\alpha = 0.2$), neither solutions could be obtained; however for $\alpha = 0.55$ we obtained $p^{\alpha} = (128.00, 350.00, 222.63, 297.37)$. This solution represent an increase around 1.06% with respect to the nominal total cost and as we can see in Figure 2.9c and Table 2.8, violations occur less frequently than when the nominal solution is selected. Finally, note that formulation based on the approximated prediction interval give us solution for $\alpha = 0.2$ and $\alpha = 0.55$. As discussed previously, the conservativeness of solutions was effectively managed through parameter α. Note that for lower values ($\alpha = 0.2$), we obtained a solution less frequently to be environmentally infeasible than for higher values ($\alpha = 0.55$). This robustness comes with a negative effect in the expected total cost. Note that (with respect to the nominal cost), we increased the total fuel cost expected from 0.44% to 1.34% by selecting a most robust solution. In any case, this increase on the cost can be seen as negligible, even more if we consider that the number of violations was significantly decreased.

Table 2.8: Comparison between deterministic and PI based robust solutions($P_D = 980$ MW/$E_{max} = 9940$ kg).

$\alpha = 0.05$ (0.1)	p_1	p_2	p_3	p_4	$F(p)$	$F(p)/F(p^{\text{nom}})$	E(p)	Viol. Freq.	Max. Viol.
p^{nom}	128.00	350.00	158.03	343.97	374.9374	1	9939.99	485	11339.88
p^{α}	-	-	-	-	-	-	-	-	-
p^{α_N}	-	-	-	-	-	-	-	-	-
$p_{\bar{f}}$	-	-	-	-	-	-	-	-	-
$\alpha = 0.20$									
p^{α}	-	-	-	-	-	-	-	-	
p^{α_N}	-	-	-	-	-	-	-	-	
$p_{\bar{f}}$	128.00	350.00	239.37	262.63	379.9516	1.0134	9162.77	60	10562.66
$\alpha = 0.55$									
p^{α}	128.00	350.00	222.63	279.37	378.9198	1.0106	9235.28	82	10635.17
p^{α_N}	-	-	-	-	-	-	-	-	-
$p_{\bar{f}}$	128.00	350.00	184.96	317.04	376.5973	1.0044	9581.96	235	10981.84

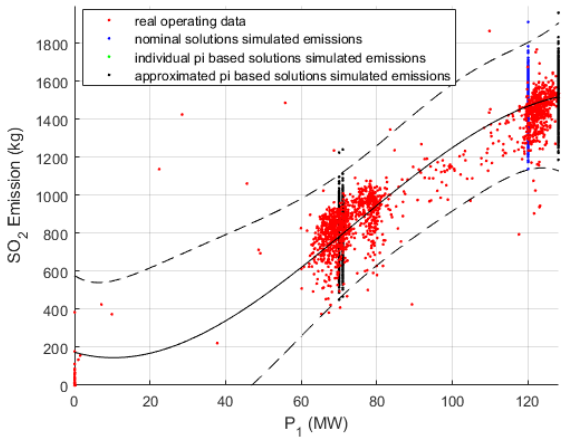
Therefore, in the above studied real system, the uncertainties in the experimentally derived models

have a significant effect on the ECED problem. When uncertainties in the coefficients of the fuel cost and emission functions are not considered, we can obtain solutions that become environmentally infeasible in some scenarios. In fact, these violations can be very frequently, depending on the parameters of the power system (specially depending on the relationship between power demand and emission level allowed). However, the quantity and maximum value of such violations can be effectively diminished by considering the respectively prediction intervals obtained through regression. In all the cases studied the increase in the total fuel cost, associated to selecting prediction interval-based solutions, instead of the nominal ones, was negligible. Finally, in Figure 2.10 we present real operating data of the studied system in three subsequent months. As we can observe, it seems that the fitted models adequately represent the real operation of the power system. Then, solutions based on prediction intervals could provide, in fact, a useful tool for manage uncertainties on the daily operation. We have also included the simulated scenarios of the four cases discussed above, in order to show the correspondence between the simulated scenarios and the real operation of the generating units.

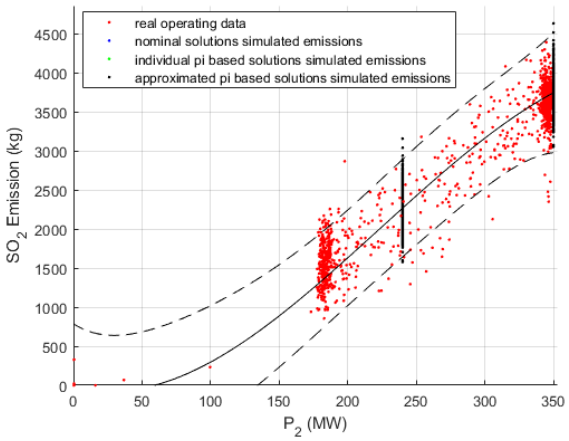
2.6 Conclusions and future works

In this chapter, an approach based on prediction intervals was proposed to solve the ECED problem. Differently to the usual deterministic approach, we incorporated to the ECED problem formulation the uncertainties associated with the experimentally derived input-output curves of thermal power plants. We analyzed the performance of the proposed approach in a real system consisting in four generating units from the Chilean electrical power network. The results obtained show that, when uncertainties are not considered, the deterministic solutions obtained can be environmentally infeasible in some scenarios; whereas the prediction interval-based solutions can effectively diminish the risk of such environmental violations, by a slight (and controllable) increase of the total fuel cost.

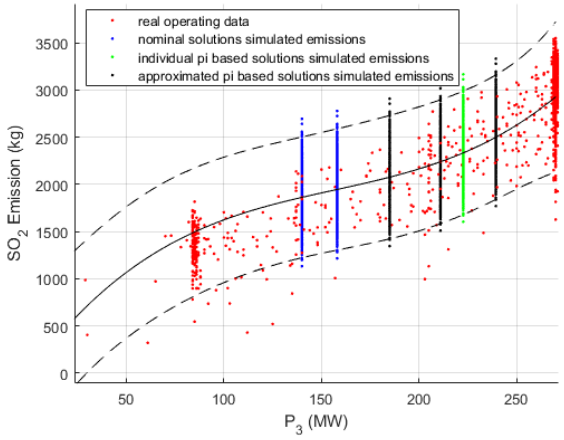
We believe that several natural extensions of this investigation exist, motivating future works. First, we need to develop efficient optimization algorithms that allow us to obtain optimal solutions to problem (2.23). This will allow us to assess the effectiveness and performance of the proposed approach in bigger electrical power networks. On another hand, statistical interval-based solutions can be studied for non-polynomial input-output curves. This can be beneficial, for example, when point-valves effect is considered in the ECED formulation.



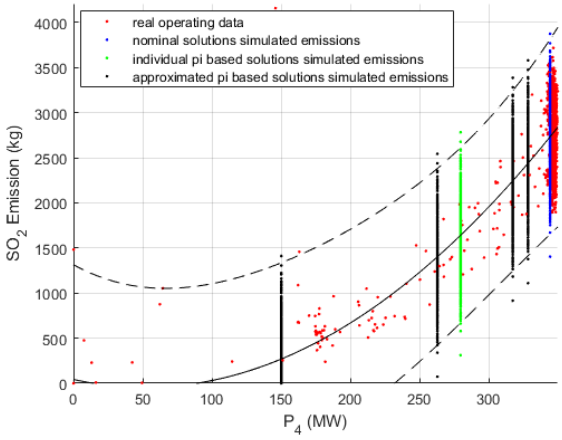
(a) Bocamina U1



(b) Bocamina U2



(c) Campiche



(d) Santa María

Figure 2.10: Simulated scenarios in all the cases analyzed.

Part II. Homogeneous optimization

Chapter 3

Non convex homogeneous optimization with applications

3.1 Introduction

In this chapter we analyze what we call as the "generalized minimum eigenvalue problem":

$$\mu_0 \doteq \min\Big\{f(x):\ g(x)=1,\ x\in C\Big\}, \tag{3.1}$$

where $C \subseteq \mathbb{R}^n$ is a (not necessarily convex) closed cone and f, g are positively homogeneous functions on C with degree p and q, respectively, such that $g(x) > 0$ for all $x \in C$, $x \neq 0$. Such a formulation, spite its simplicity, encompasses several important problems in many different areas, as we briefly describe below. One method to approximate μ_0 is via the optimal value of a "dual problem", which is not uniquely determined. We propose the following (Lagrangian) dual problem:

$$\nu \doteq \sup_{\lambda\in\mathbb{R}} \inf_{x\in C} \Big\{f(x)+\lambda(g(x)-1)\Big\}. \tag{3.2}$$

Obviously ν is always a lower bound for μ_0 (weak duality). It worth-while to mention that the dual problem admits the unique solution $-\mu_0\dfrac{p}{q}$ (Theorem 3.1).

One of the main purposes of this work is to discuss the validity of the equality $\mu_0 = \nu$ (zero duality gap), and the strong duality property, which means that $\mu_0 = \nu$ and problem (3.2) has solution; among other important issues including KKT optimality.

In many practical models, C is described by inequalities and/or equalities, and so, one can talk about standard dual problem, whose associated Lagrangian involves all the constraint functions, and so, the infimum in (3.2) is taken over $\mathbb{R}^n$. With respect to this standard dual problem, we will see that zero duality gap, or strong duality property occurs under very strong assumptions.

More precisely, we establish various characterizations for the fulfillment of strong duality property (Theorem 3.1). In particular, it holds if, and only if $(g, f)(C)+\mathbb{R}_+(0,1)$ is convex. This is satisfied if either $\mu_0 = 0$ or $p = q$ (this holds for some of the models below), and in the latter situation, the copositive formulation holds (Theorem 3.1 and Corollary 3.1):

$$\mu_0 = \nu = \sup\Big\{\lambda :\ f - \lambda g \text{ is copositive on } C\Big\},$$
$$= \inf\Big\{f(x) - \mu_0(g(x)-1) :\ x \in C\Big\},$$

where by copositivity on C of any function h, we mean that $h(x) \geq 0$ for all $x \in C$.
Furthermore, we introduce the notion of L(agrange)-eigenvalue (see Subsection 3.4.2) as a Lagrange multiplier associated to the existence of a KKT-point (different from the zero vector) suitably defined, which is related to a necessary optimality condition for the problem

$$\inf\Big\{f(x) - \mu_0\frac{p}{q}(g(x)-1) :\ x \in C\Big\}.$$

In some sense, our approach provides a variational scheme to the analysis of eigenvalues to certain mappings; in particular, to real symmetric tensors (see Model 2). For other perspective we refer to [74].
Before describing some of the main models where our approach applies, we list just a few concrete applications. For instance, a class of quadratic homogeneous optimization problems arises in telecommunications and robust control as reported in [99, 131]; minimum eigenvalue of a symmetric matrix; minimization of a homogeneous polynomial over spheres or hyperspheres (for instance [89, 108], giving rise to the sum of squares (SOS) relaxation as an aproximation method); several important classes of quadratic programming problems lying in matrix theory; special relativity [51]; trust region problems [61, 141] (for recent advances on this matter, we refer [162]).

Model 1: Quadratic fractional optimization problems with two quadratic constraints
This problem, to be analyzed in detail in Section 3.6, takes the form:

$$\mu_{qf} \doteq \min\Big\{\frac{w^\top A w + a^\top x + \alpha}{w^\top B w + b^\top w + \beta} :\ w^\top B_1 w + b_1^\top w + \beta_1 \leq 0,\ w^\top w - \delta \leq 0\Big\}, \tag{3.3}$$

where $\delta > 0$; a, b, b_1 are in $\mathbb{R}^n$; $\alpha, \beta, \beta_1 \in \mathbb{R}$; and A, B, B_1 are (real) symmetric matrices of order n satisfying $w^\top B w + b^\top w + \beta > 0$ for all feasible w for (3.3). By using the generalized Charnes-Cooper transformation, problem (3.3) reduces to one of the form (3.1) with $p = q = 2$, satisfying the required assumptions. In particular, a copositive reformulation can be obtained (see Corollary 3.1), besides the validity of strong duality property (Theorem 3.1). This problem was studied in [52] under the SDP relaxation approach, and so yielding tightness, which in turns implies that standard strong duality holds. Our approach considers situations where standard strong duality may fail. In addition, we establish a new necessary and sufficient second-order optimality conditions (Theorem 3.11). This will allow us, under very mild assumptions on the reference point, to identify optimality by means of a first-order optimality condition (Theorem 3.12). Therefore, the problem is solvable by using the notion of (Moore-Penrose) pseudoinverse.

Model 2: Tensors eigenvalues analysis
The analysis will be developed in detail in Section 3.7 and it refers to the problem:

$$\mu_k \doteq \min\Big\{\mathcal{A}x^m :\ \|x\|_k^m = 1,\ x \in C\Big\}, \tag{3.4}$$

where $\mathcal{A}$ is an m-order n-dimensional real symmetric tensor, and so $\mathcal{A}x^m$ defines a homogeneous polynomial of degree m; $\|x\|_k$ denotes the l^k-norm, and C is a closed convex cone in $\mathbb{R}^n$. It will be

showed that μ_k is the least L(agrange)-eigenvalue (to be introduced in Subsection 3.4.2) associated to problem (3.4). By particularizing $k = 2$ or $k = m$ and either $C = \mathbb{R}^n$ or $C = \mathbb{R}^n_+$, we recover the notion of Z-eigenvalue (eigenvector) or H-eigenvalue (eigenvector), see [92, 122, 129, 137], among others. Hence, we provide a unified approach in tensor analysis, and produce new results about the copositivity meaning in this context (Proposition 3.8).

Model 3: Approaching linear complementarity problems
A quadratic programming approach to linear complementary problems (LCP) is to consider the bilinear program (see, for instance [8]):

$$\min\{z^\top w : \ -Mz + w = q, \ z \geq 0, \ w \geq 0\}.$$

This problem can be written, in a equivalent way as:

$$\min\{x^\top Ax : \ Hx = q, \ x \geq 0\}, \tag{3.5}$$

where:

$$x = \begin{pmatrix} z \\ w \end{pmatrix}, \qquad A = \left(\begin{array}{c|c} 0 & \frac{1}{2}\, I \\ \hline \frac{1}{2}\, I & 0 \end{array}\right), \qquad H = \begin{pmatrix} -M \mid I \end{pmatrix},$$

and 0, I stand for the null and identity matrices of order $n \times n$, respectively, and A is indefinite but copositive on $\mathbb{R}^{2n}_+$, i.e., $x^\top Ax \geq 0$ for all $x \in \mathbb{R}^{2n}_+$.
We consider an example from [35] to illustrate how to re-formulate (3.5) in the form (3.1). In such an example, we have:

$$M = \begin{pmatrix} 0 & 0 & 0 \\ 1 & 0 & 0 \\ -1 & -1 & -1 \end{pmatrix}, \qquad q = \begin{pmatrix} 1 \\ -1 \\ 3 \end{pmatrix}.$$

By setting $f(x) \doteq x^\top Ax$, $g(x) \doteq \dfrac{1}{3}(x_1 + x_2 + x_3 + x_6)$, $g_1(x) \doteq -x_1 - x_2 - x_3 + 3x_4 - x_6$, $g_2(x) = 4x_1 + x_2 + x_3 + 3x_5 + x_6$, and choosing as C to be the closed convex cone $C \doteq \{x \in \mathbb{R}^6 : x \geq 0, g_1(x) = 0, g_2(x) = 0\}$, we have that f and g are positively homogeneous functions of degrees $p = 2$ and $q = 1$, respectively. Moreover, $g(x) > 0$ for all $x \in C$, $x \neq 0$. Certainly, f is nonconvex and $f(x) \geq 0$ for all $x \geq 0$. We refer to [56] for a great account about linear complementarity problems.

Model 4: Extensions of the standard quadratic and portfolio optimization problems
We consider now two problems with the following general structure:

$$\min\Big\{f(x) : \ e^\top x = 1, \ x \in C\Big\},$$

where $C \subseteq \mathbb{R}^n$ is a pointed, closed, convex cone having non-empty interior, and $e \in \operatorname{int} C^*$. Here, C^* is the non-negative polar cone of C.
For $f(x) = \dfrac{1}{2}x^\top Ax$, with A a real symmetric matrix of order n, problem:

$$\mu_q \doteq \min\Big\{\frac{1}{2}x^\top Ax : \ e^\top x = 1, \ x \in C\Big\}, \tag{3.6}$$

is an extension of the standard quadratic optimization problem (StQOP). This model generalize the problem introduced by Bomze in [22] (where $C = \mathbb{R}^n_+$, $e = \mathbb{1} \doteq (1, \ldots, 1)$), which models: quadratic allocation problems [77]; (classical mean-variance) portfolio optimization problems [97, 98]; the maximum weight clique problem [66, 105]; the indefinite quadratic knapsacks problem [114], see also [99,131], among others. Due to the structure of the feasible set, it is not restrictive to consider homogeneous functions, since:

$$\frac{1}{2}x^\top Ax + a^\top x = \frac{1}{2}x^\top (A + ae^\top + ea^\top)x.$$

The StQOP was introduced in [22] and further developed in [23–27] and references therein.
According to our Theorem 3.1, problem (3.6) has strong duality if, and only if A is copositive, since $p = 2 > q = 1$. On the other hand, we can find a copositive formulation of (3.6), since it can be written equivalently as:

$$\mu_q \doteq \min\Big\{\frac{1}{2}x^\top Ax : \ x^\top ee^\top x = 1, \ x \in C\Big\},$$

whose dual problem is:

$$\nu_q \doteq \sup_{\lambda\in\mathbb{R}} \inf_{x\in C} \Big\{\frac{1}{2}x^\top Ax + \lambda(x^\top ee^\top x - 1)\Big\}.$$

By particularizing Corollary 3.1 ($p = q = 2$), one gets:

$$\mu_q = \nu_q = \sup_{\lambda\in\mathbb{R}}\Big\{ -\lambda : \ A + 2\lambda ee^\top \text{ is copositive on } C\Big\}.$$

The case $C = \mathbb{R}^n_+$, allows us to introduce the standard dual problem, as we said at the beginning. In this situation, standard strong duality is satisfied if, and only if A is positive semidefinite; and therefore, in practice such a dual is useless.

On another hand, a different model to the mean-variance portafolio optimization problem (3.6), is that known as the "standard deviation premium" considered in [87], where the variance is replaced by the standard deviation, that is, f takes the form $f(x) = a^\top x + \rho\sqrt{x^\top Vx}$, so the problem is:

$$\mu_p \doteq \min\Big\{a^\top x + \rho\sqrt{x^\top Vx} : \ e^\top x = 1, \ x \in C\Big\}.$$

We refer to [88] for a further discussion. Here, V is only required to be copositive on C. Since $p = q = 1$, strong duality holds because of Theorem 3.1, and by Corollary 3.1, the copositive formulation ($g(x) = e^\top x$) is:

$$\mu_p = \max\{-\lambda : \ f + \lambda g \text{ is copositive on } C\}.$$

If, instead, one considers the equivalent constraint $g(x) \doteq x^\top ee^\top x = 1$, then $p = 1 < q = 2$, and so by the same theorem, strong duality is satisfied provided $f(\bar{x}) < 0$ for some $\bar{x} \in C$.
Several other problems, after some mathematical manipulations like in Model 3, can also be formulated as in (3.1).

This chapter is organized as follows. Section 3.2 serves to introduce some basic definitions and preliminaries, as well as to revisit the Lagrangian duality scheme. Section 3.3 establishes various

new characterizations of the validity of: strong duality for (3.1) (revealing convexity as Theorem 3.1 shows); the S-lemma (Lemma 3.2), a copositive formulation for (3.1) when $p = q$, as Corollary 3.1 shows. In Section 3.4, it is discussed: zero-order optimality conditions; KKT optimality and L-eigenvalues; the case when C is expressed by two quadratic forms, yielding new necessary optimality conditions under strong duality; and when $p = q = 2$, those conditions become also sufficient. Section 3.5 is devoted to compare our approach with the SDP relaxation scheme: this is carried out for two examples discussed in [107]. Section 3.6 fully analyzes a class of quadratic fractional optimization problems with two quadratic constraints. Finally, in Section 3.7 the case of a real m-order n-dimensional supersymmetric tensor is discussed. In particular, some relationships linking our notion of L-eigenvalue and those of Z-eigenvalue or H-eigenvalue, are presented.

3.2 Some notation, basic definition and preliminaries

Throughout this chapter, we will work on a finite dimensional space, say $\mathbb{R}^m$. Given any nonempty set M in $\mathbb{R}^m$, its closure, topological interior, convex hull, closed convex hull, are denoted, respectively, by $\overline{M}$, int M, co M, $\overline{\text{co}}\ M$. In addition, by aff M, span M, ri M and bd M we denote the affine set of M, span of M, relative interior of M and the boundary of M, respectively. Moreover, cone M is the smallest cone containing M, i.e., cone $M = \bigcup_{t\geq 0} tM$. The polar cone of M is defined by:

$$M^* \doteq \{z \in \mathbb{R}^m :\ \langle z, y\rangle \geq 0\ \ \forall\ y \in M\}.$$

Here, $\langle z, y\rangle = z^\top y$ stands for the scalar product between two vectors z and y in $\mathbb{R}^m$, where $z^\top$ means the transpose of the vector z, which is considered a column vector. More generally, if A is a real matrix in $\mathbb{R}^{m\times n}$, $A^\top$ is the transpose of A belonging to $\mathbb{R}^{n\times m}$.
Let $h : \mathbb{R}^m \to \mathbb{R} \cup \{\pm\infty\}$, $\overline{h}$ and $\overline{\text{co}}\ h$ stand for the greatest lower semicontinuous function not larger than h and for the greatest convex and lower semicontinuous function not larger than h, respectively. Just for convenience, we need the following definition of epigraph of a function: epi $h \doteq \{(y, t) \in \mathbb{R}^m \times \mathbb{R} :\ h(y) \leq t\}$. It is known that:

$$\text{epi}\ \overline{h} = \overline{\text{epi}\ h};\ \ \overline{\text{co}}(\text{epi}\ h) = \text{epi}\ \overline{\text{co}}\ h.$$

Moreover,

$$\overline{\text{co}}\ h(y) > -\infty\ \ \forall\ y \in \mathbb{R}^m \Longrightarrow \overline{\text{co}}\ h(y) = h^{**}(y)\ \ \forall\ y \in \mathbb{R}^m,$$

where $h^{**} = (h^*)^*$ is the bipolar or biconjugate of h, that is, the conjugate (or polar) of h^* defined by:

$$h^*(z) \doteq \sup_{y\in\mathbb{R}^m} \{\langle z, y\rangle - h(y)\}.$$

In addition, δ_M stands for the indicator function of the set M, defined by 0 on M, and $+\infty$ on the complementary of M.
There are examples showing the assumption $\overline{\text{co}}\ h(y) > -\infty$ for all $y \in \mathbb{R}^m$ is necessary to get the equality $h^{**} = \overline{\text{co}}\ h$. In general we have $h^{**} \leq \overline{\text{co}}\ h \leq h$. For details see [124].
In case $h : \mathbb{R}^m \to \mathbb{R} \cup \{+\infty\}$, the subdifferential of h at $\bar{y} \in \mathbb{R}^m$ is denoted by:

$$\partial h(\bar{y}) \doteq \{\xi \in \mathbb{R}^m : \ h(y) \geq h(\bar{y}) + \langle \xi, y - \bar{y} \rangle, \ \forall \ y \in \mathbb{R}^m\},$$

if $\bar{y} \in \text{dom } h$, and $\partial h(\bar{y}) = \emptyset$ elsewhere.
In the subsequent sections, we set $\mathbb{R}_+ \doteq [0, +\infty[$; $\mathbb{R}_{++} \doteq]0, +\infty[$. Given a vector $a \in \mathbb{R}^m \setminus \{0\}$, $\mathbb{R}_+ a$ stands for the ray starting from the origin and direction a; and $a^\perp$ is the orthogonal subspace to a, which is a hyperplane.
In the remaining part of this section, we present a duality scheme for a minimization problem under one single equality constraint and a geometric constraint set, mainly taken from Section 3 in [54]. Let $f, g_0 : \mathbb{R}^n \to \mathbb{R}$ be any finite-valued functions, and let $C \subseteq \mathbb{R}^n$ be any nonempty set. Let us consider the problem:

$$\mu \doteq \inf\{f(x) : \quad g_0(x) = 0, \ x \in C\}, \tag{3.7}$$

whose (Lagrangian) dual problem is defined by:

$$\nu \doteq \sup_{\lambda \in \mathbb{R}} \inf_{x \in C} [f(x) + \lambda g_0(x)]. \tag{3.8}$$

We say that there is no duality gap, or the duality gap is zero, between (3.7) and (3.8) if $\nu = \mu$. It is said that (3.7) has the strong duality property with respect to (3.8), or simply that strong duality holds for (3.7) if $\mu = \nu$ and problem (3.8) admits a solution. One infers immediately that $\nu \leq \mu$. Thus, if $\mu = -\infty$ then there is no duality gap, and we conclude that any element in $\mathbb{R}$ is a solution for the problem (3.8). Hence, we always have strong duality for (3.7) whenever $\mu = -\infty$. So, we suppose from now on that $\mu \in \mathbb{R}$, which means, in particular, that the feasible set to (3.7) is nonempty.
Set $F(x) \doteq \Big(g_0(x), f(x)\Big)$. Notice that $F = (f, g_0)$ was used in [55] instead. Assuming that $\mu \in \mathbb{R}$, we obtain:

$$(F(C) - \mu(0,1)) \cap -(\{0\} \times \mathbb{R}_{++}) = \emptyset. \tag{3.9}$$

We will show, next, that strong duality can be characterized by reinforcing (3.9).
The optimal value function $\psi : \mathbb{R} \to \mathbb{R} \cup \{\pm\infty\}$ to problem (3.7) is defined by:

$$\psi(a) = \begin{cases} \inf\{f(x) : \ x \in K(a)\} & \text{if} \quad K(a) \neq \emptyset; \\ +\infty \qquad\quad \text{otherwise,} \end{cases}$$

where:

$$K(a) \doteq \{x \in C : \ g_0(x) = a\}.$$

Notice that $K = K(0)$, and $K(a) \neq \emptyset$ if and only if $a \in g_0(C)$, that is,

$$\text{dom } \psi \doteq \{a \in \mathbb{R} : \ \psi(a) < +\infty\} = g_0(C).$$

The sets:

$$\mathcal{F} \doteq F(C) + \mathbb{R}_+(0,1), \ \mathcal{E}_\rho \doteq \mathcal{F} - \rho(0,1) \quad (\rho \in \mathbb{R}),$$

will play an important role in our analysis.

Remark 3.1. (*a*) *By definition, strong duality holds if and only if there exists* $\lambda_0 \in \mathbb{R}$ *such that*

$$f(x) + \lambda_0 g_0(x) \geq \mu, \ \forall \ x \in C,$$

or equivalently, $\mathcal{L}_{SD} \neq \emptyset$*, where*

$$\mathcal{L}_{SD} \doteq \{\lambda_0 \in \mathbb{R} : \ (\lambda_0, 1) \in (\mathcal{E}_\mu)^*\}.$$

Hence, $\mathcal{L}_{SD} \subseteq \mathcal{S}_D$ *with* $\mathcal{S}_D$ *being the solution set to the dual problem* (3.8)*. Moreover,* $\mathcal{L}_{SD} = \mathcal{S}_D$ *whenever zero duality gap holds.*
(*b*) *The following chain of inclusions shows useful and well-known properties of the optimal value function* ψ *(see [55] for instance):*
$F(C) + \mathbb{R}_+(0,1) \subseteq \text{epi } \psi \subseteq \overline{F(C) + \mathbb{R}_+(0,1)}$*. Consequently,*

$$\overline{\mathcal{E}_\mu} = \overline{\text{epi } \psi} - \mu(0,1) = \text{epi } \overline{\psi} - \mu(0,1); \quad \overline{\text{co}} \ \mathcal{E}_\mu = \overline{\text{co}}(\text{epi } \psi) - \mu(0,1) = \text{epi}(\overline{\text{co}} \ \psi) - \mu(0,1).$$

The next result states various equivalences, of topological or geometric nature, to the validity of strong duality.

Proposition 3.1. (*[55, Theorem 4.2]*) *Assume that* $\mu = \psi(0)$ *is finite. The following assertions are equivalent:*

(*a*) *Strong duality holds for* (3.7)*;*

(*b*) $\overline{\text{cone}}(\text{co } \mathcal{E}_\mu) \cap (-\{0\} \times \mathbb{R}_{++}) = \emptyset$;

(*c*) $\partial\psi(0) \neq \emptyset$*;*

(*d*) $\overline{\text{cone}}(\mathcal{E}_\mu) \cap (-\{0\} \times \mathbb{R}_{++}) = \emptyset$ *and* $\overline{\text{cone}}(\mathcal{E}_\mu)$ *is convex.*

Hence, under any of the above conditions, one gets

$$\partial\psi(0) = \{-\lambda_0 \in \mathbb{R} : \ \lambda_0 \in \mathcal{S}_D\}.$$

We must point out that a more detailed description of the disjointness appearing in (*d*), is presented in [28, Theorem 3]. On the other hand, one can see from the previous proposition that the convexity of $\overline{\text{cone}}(\mathcal{E}_\mu)$ arises in a natural way under strong duality no matter the functions f and g are. The equivalence between the validity of strong duality and the convexity of $\overline{\text{cone}}(F(C) + \mathbb{R}_+(0,1) - \mu(0,1))$ was proved in [28,55] under a Slater-type condition. In case $C = \mathbb{R}^n$ and f, g are quadratic functions, the authors in [59] established (under a Slater condition) that strong duality holds if and only if $F(\mathbb{R}^n) + \mathbb{R}_+(0,1)$ is convex. When C is a pointed closed convex cone, and f is a quadratic form and g linear (see Model 4), such an equivalence with the convexity of $F(C) + \mathbb{R}_+(0,1)$, was established in [54].
We have to mention that in case we have more than one constraint, one may proceed by including all the constraints, except one, in the geometric constraint set C. Among the recent results about the convexity of images of quadratic functions, we mention [17, 54, 59, 82, 157].

3.3 Formulation of the problem: characterizing strong duality and S-lemma; copositive reformulation

Let us go back to our original problem formulated in Section 3.1:

$$\mu_0 \doteq \min\Big\{f(x):\ g(x)=1,\ x\in C\Big\}, \tag{3.10}$$

whose feasible set is denoted by K (which is supposed to be nonempty), and where $C\subseteq\mathbb{R}^n$ is a closed cone, and $f,g:\mathbb{R}^n\to\mathbb{R}$ satisfy the following assumption:
Assumption (A):
Let p,q be positive real numbers. The functions f and g are lower semicontinuous (lsc, in short) such that:

(i) $f(tx)=t^pf(x)$ for all $t>0$ and all $x\in C$;

(ii) $g(tx)=t^qg(x)$ for all $t>0$ and all $x\in C$;

(iii) $g(x)>0$ for all $x\in C$, $x\neq 0$; as a consequence, $x\in C$, $g(x)=0$ if, and only if $x=0$.

Some remarks are in order.

Remark 3.2. *Let C be a closed cone.*
(a) Let $h:\mathbb{R}^n\to\mathbb{R}\cup\{+\infty\}$ be a lsc function and positively homogeneous with degree p. Then, $h(0)=0$; the sets $C\cap\operatorname{dom} h$, $C\setminus\operatorname{dom} h$ are cones. Moreover, in case h is differentiable at $x\in C\cap\operatorname{dom} h$, the so-called Euler identity holds:

$$\nabla h(x)^\top x = ph(x).$$

In addition, ∇h is positively homogeneous with degree $p-1$.
(b) Notice that under (iii) and lsc on g, the set $\{x\in C:\ g(x)\leq\gamma\}$ is bounded for all $\gamma\geq 0$.
(c) In most applications C is, in addition, convex and pointed, and $g(x)=e^\top x$ with $e\in\operatorname{int} C^*$; thus g satisfies (iii). Another useful specialization is $g(x)=(\|x\|_m)^m\doteq|x_1|^m+|x_2|^m+\cdots+|x_m|^m$ and $C=\mathbb{R}^m_+$.

The fact that the cone C is not necessarily convex makes our model to be very versatil, as described at the introduction section. On the other hand, by virtue of Assumption (A) and since K is nonempty, problem (3.10) always fulfills the Slater condition: there exist $x_1,x_2\in C$ such that $g(x_1)<1<g(x_2)$.

It is said that $h:\mathbb{R}^n\to\mathbb{R}\cup\{+\infty\}$ is copositive on a cone $P\subseteq\mathbb{R}^n$ if $h(x)\geq 0$ for all $x\in P$; it is strictly copositive on P if $h(x)>0$ for all $x\in P$, $x\neq 0$. By extension, a (real) symmetric matrix A is said to be (resp. strictly) copositive on P, if the function $h(x)=x^\top Ax\geq 0$ (resp. >0) for all $x\in P$ (resp. for all $x\in P, x\neq 0$).
By assumption, one gets:

$$C=\bigcup_{t\geq 0} tK,$$

with K being a compact set. Thus, it is easy to check that:

- $\mu_0 \geq 0 \Longleftrightarrow f$ is copositive on C;
- $\mu_0 = 0 \Longleftrightarrow f$ is copositive but not strictly copositive on C;
- $\mu_0 > 0 \Longleftrightarrow f$ is strictly copositive on C.

The following lemma shows some intrinsic properties of problem (3.10), and relationships with the problems:

$$\mu_+ \doteq \inf\Big\{f(x):\ g(x) \geq 1,\ x \in C\Big\};\ \mu_- \doteq \inf\Big\{f(x):\ g(x) \leq 1,\ x \in C\Big\}.$$

Set $K_+ \doteq \{x \in C:\ g(x) \geq 1\}$ and $K_- \doteq \{x \in C:\ g(x) \leq 1\}$.

Lemma 3.1. *Let C be a closed cone, and f, g satisfy Assumption* (A)*. The following assertions hold.*

(a) *Assume that $+\infty > \mu_0 \geq 0$. Then,*

($a1$) $x \in C,\ f(x) < \mu_0 \Longrightarrow g(x) < 1$.

($a2$) $\mu_0 = \mu_+$ *and* $\operatorname*{argmin}_K f \subseteq \operatorname*{argmin}_{K_+} f$.

($a3$) *If* $\mu_0 > 0$, *then* $\operatorname*{argmin}_K f = \operatorname*{argmin}_{K_+} f$, *and* $[x \in C,\ f(x) = \mu_0 \Longrightarrow g(x) \leq 1]$.

(b) *Assume that* $\mu_0 \leq 0$. *Then,*

($b1$) $x \in C,\ f(x) < \mu_0 \Longrightarrow g(x) > 1$.

($b2$) $\mu_0 = \mu_-$ *and* $\operatorname*{argmin}_K f \subseteq \operatorname*{argmin}_{K_-} f$.

($b3$) *If* $\mu_0 < 0$, $\operatorname*{argmin}_K f = \operatorname*{argmin}_{K_-} f$, *and* $[x \in C,\ f(x) = \mu_0 \Longrightarrow g(x) \geq 1]$.

Proof. We only prove (a), being the other entirely similar.
($a1$): Take any $x \in C$ such that $f(x) < \mu_0$. Obviously the implication holds if $x = 0$. Suppose now that $g(x) > 1$, and write $x = ty$ for some $t > 0$ and $y \in K$. Thus $t^q > 1$ and so $t > 1$. Moreover, we obtain $\mu_0 > f(x) = t^p f(y) \geq t^p \mu_0$, which implies $\mu_0 < 0$, and this contradicts our assumption.
($a2$): Clearly $\mu_+ \leq \mu_0$. Suppose that $\mu_+ < \mu_0$ and choose $y \in K_+$ satisfying $f(y) < \mu_0$. By ($a1$), $g(y) < 1$, a contradiction.
($a3$) The first part follows easily, and the second one is similar to ($a2$). □

Denote, given $a \in \mathbb{R}$,

$$K(a) \doteq \{x \in C:\ g(x) = 1 + a\}.$$

It is worth noticing that a complete study in the case when $p = 2$, $q = 1$ was carried out in [54], including some necessary or sufficient conditions for local or global optimality.
The following proposition, whose proof is straightforward, collects some basic facts on the optimal value function:

$$\psi(a) \doteq \min_{x \in K(a)} f(x).$$

In order to make compatible problem (3.10) with that of (3.7) we introduce the function $g_0(x) \doteq g(x) - 1$, and so set $F \doteq (g_0, f)$.

Proposition 3.2. *Let C be a closed cone, and f, g satisfy Assumption* (A)*. The following assertions hold.*

(*a*) $K(a) \neq \emptyset$ *if, and only if* $a \geq -1$.

(*b*) *Let* $a > -1$. *Then,* $x \in K$ *if, and only if* $(a+1)^{1/q}x \in K(a)$.

(*c*) *The optimal value function is given by:*

$$\psi(a) = \begin{cases} \mu_0(1+a)^{p/q} & \text{if} \quad a \geq -1; \\ +\infty & \text{if} \quad a < -1. \end{cases}$$

(*d*) *Assume that f and g are continuous. Then $F(C)$ and $F(C)+\mathbb{R}_+(0,1)$ are closed, so* epi $\psi = F(C) + \mathbb{R}_+(0,1)$.

Proof. (*a*): It is obvious.
(*b*): Let $x \in K$. Then $g((a+1)^{1/q}x) = (a+1)g(x) = a+1$. Thus $(a+1)^{1/q}x \in K(a)$, and so, by symmetry, the result follows.
(*c*): It is a consequence of (*b*).
(*d*): We check the closedness of $F(C) + \mathbb{R}_+(0,1)$. The same argument also shows that $F(C)$ is closed. Let $(a,r) \in \overline{F(C)+\mathbb{R}_+(0,1)}$. Then, there exist sequences $x_k \in C$, $q_k \geq 0$ satisfying $f(x_k) + q_k \to r$ and $g(x_k) - 1 \to a$. By assumption, the second relation implies the boundedness of $\|x_k\|$. Thus, up to a subsequence, $x_k \to \bar{x} \in C$, implying that $q_k = f(x_k) + q_k - f(x_k) \to r - f(\bar{x})$. Setting $q \doteq r - f(\bar{x})$, we get $q \geq 0$, and so $(a,r) = (g(\bar{x}) - 1, f(\bar{x}) + q) \in F(C) + \mathbb{R}_+(0,1)$. The last part follows from (*b*) of Remark 3.1 when it applies to $g_0(x) = g(x) - 1$. □

By Proposition 3.2, it is not difficult to prove that:

$$\overline{\text{cone}}(F(C) + \mathbb{R}_+(0,1) - \mu_0(0,1))$$

$$= \begin{cases} \{(u,v) \in \mathbb{R}^2 : \ v \geq \dfrac{\mu_0 p}{q} u\} \text{ if } [0 < q \leq p, \ \mu_0 \geq 0] \ \text{ or } [0 < p \leq q, \ \mu_0 \leq 0]; \\ \{(u,v) \in \mathbb{R}^2 : \ v \geq \mu_0 u\} \cup (\mathbb{R}_+ \times \mathbb{R}) \text{ if } 0 < q < p, \ \mu_0 < 0; \\ \{(u,v) \in \mathbb{R}^2 : \ v \geq \mu_0 u\} \cup (\mathbb{R} \times \mathbb{R}_+) \text{ if } 0 < p < q, \ \mu_0 > 0. \end{cases} \tag{3.11}$$

We now establish how the fulfillment of strong duality property reveals the hidden convexity of some image set. To the best of our knowledge, this result is new and the first one (in a non-local sense) for functions beyond the quadratic world.

Theorem 3.1. *Let C be a closed cone, and f, g satisfy Assumption* (A)*. Then, the following are equivalent:*

(*a*) *strong duality holds.*

(*b*) $F(C) + \mathbb{R}_+(0,1)$ *is convex.*

(*c*) *Exactly one of the following assertions is satisfied:*

(*c*1) f *is copositive but not strictly copositive on* C $(\mu_0 = 0)$.

(*c*2) f *is strictly copositive on* C $(\mu_0 > 0)$ *and* $p \geq q > 0$.

(*c*3) f *is not copositive on* C $(\mu_0 < 0)$ *and* $q \geq p > 0$.

Consequently, under any of conditions (*a*), (*b*) *or* (*c*), *the unique solution to the dual problem* (3.2) *is* $-\dfrac{p}{q}\mu_0$, *and so*

$$\mu_0 = \inf_{x \in C}[f(x) - \mu_0 \frac{p}{q}(g(x) - 1)]. \tag{3.12}$$

Proof. The equivalence between (*b*) and (*c*) follows from (*c*) of Proposition 3.2. The equivalence between (*b*) and (*a*) is a consequence of (3.11) and Proposition 3.1. The remaining part also follows from Proposition 3.1. □

One of the interpretations of Theorem 3.1 follows. In case $0 < q \leq p$, Theorem 3.1 characterizes the strict copositivity on C of every function f that is positively homogeneous with degree p on C, by means of the convexity of $(g, f)(C) + \mathbb{R}_+(0, 1)$ for some (any) function g positively homogeneous with degree q on C.

From Theorem 3.1, we realize that when $p = q$ it is possible to establish a copositive reformulation of the dual problem, and so of the primal one. Such a formulation extends Lemma 3.2 and (main) Theorem 3.5 in [121], which considers the quadratic (homogeneous) case with $C = \mathbb{R}^n_+$.

Corollary 3.1. (*Copositive formulation*) *Let* $p = q > 0$. *Assume that* f *and* g *satisfy Assumption* (A). *Then,*

$$\mu_0 = \nu = \max\Big\{\lambda : \ f - \lambda g \text{ is copositive on } C\Big\}, \quad \textit{and}$$

- $f - \mu_0 g$ *is copositive but not strictly copositive on* C; *that is, there exists* $\bar{x} \in C, \bar{x} \neq 0$, *such that*
$$\mu_0 = \min_{\substack{x \in C \\ x \neq 0}} \frac{f(x)}{g(x)} = \frac{f(\bar{x})}{g(\bar{x})}.$$
- $\forall\ \gamma < \mu_0$, $f - \gamma g$ *is strictly copositive on* C;
- $\forall\ \gamma > \mu_0$, $f - \gamma g$ *is not copositive on* C.

Proof. By the previous theorem we get strong duality, and so

$$\begin{aligned}
\mu_0 = \nu &= \sup_{\lambda \in \mathbb{R}} \inf_{x \in C}[f(x) - \lambda(g(x) - 1)] = \sup_{\lambda \in \mathbb{R}} \Big\{\lambda + \inf_{x \in C}[f(x) - \lambda g(x)]\Big\} \\
&= \sup_{\lambda \in \mathbb{R}} \Big\{\lambda : \ \inf_{x \in C}[f(x) - \lambda g(x)] = 0\Big\} \\
&= \max\Big\{\lambda : \ f - \lambda g \text{ is copositive on } C\Big\}.
\end{aligned}$$

The remaining assertions are straightforward. □

We utilize Corollary 3.1 and the bisection algorithm to propose an algorithm to determine μ_0 or an approximation of it once a tolerance $\varepsilon > 0$ is prescribed. Here, f, g and C are as above.

Algorithm

1. Select γ_- and γ_+ such that $f - \gamma_- g$ is strictly copositive (so $\gamma_- < \mu_0$) and $f - \gamma_+ g$ is not strictly copositive on C (so $\mu_0 < \gamma_+$).

2. Set $\gamma \doteq (\gamma_- + \gamma_+)/2$.

3. If $f - \gamma g$ is strictly copositive, set $\gamma_- \doteq \gamma$. Otherwise, $\gamma_+ \doteq \gamma$.

4. If $f - \gamma_+ g$ is copositive or if $\gamma_+ - \gamma_- < \varepsilon$, then $\mu_0 = \gamma_+$ and stop. Otherwise, go to step **2**.

This algorithm, when f and g are quadratic forms and $C = \mathbb{R}^n_+$, was discussed in more detail in [121].

We are now ready to establish the S-lemma suitable for the problem (3.10). Such a lemma, which will be formulated in its strict version, asks for the equivalence between (3.13) and (3.14):

$$x \in C,\ g(x) - 1 = 0 \Longrightarrow f(x) > 0; \tag{3.13}$$

$$\exists\ \lambda \neq 0,\ f(x) + \lambda(g(x) - 1) > 0\ \forall\ x \in C. \tag{3.14}$$

Results of this kind goes back to the works by Yakubovich in [160, 161], who considered f, g to be quadratic forms. To be more precise, his version reads as follows: the next two statements are equivalent provided there exists $\bar{x}$ satisfying $\bar{x}^\top B \bar{x} < 0$,

$$x \in \mathbb{R}^n, x^\top B x \leq 0 \Longrightarrow x^\top A x \geq 0.$$

$$\exists\ \lambda \in \mathbb{R} : A + \lambda B \text{ is positive semidefinite.}$$

Its inhomogeneous version was studied in [59,157]. A further development when C is a convex cone is presented in [54]. A nice survey (until 2007) on the S-lemma in the quadratic word is [120]. More recent results may be found in [60, 157].

Lemma 3.2. (*S-lemma*) *Let C be a closed cone, g satisfies (ii) and (i) of Assumption* (A). *The following assertions are equivalent:*

(a) *for all lsc function f and positively homogeneous with degree $p > 0$ on C, one has*

$$[x \in C,\ g(x) = 1 \Longrightarrow f(x) > 0] \Longleftrightarrow \exists\ \lambda \neq 0,\ f(x) + \lambda(g(x) - 1) > 0\ \forall\ x \in C.$$

(b) $p \geq q$.

Proof. (a) $\Rightarrow$ (b): Take $f(x) \doteq |x_1|^p + \cdots + |x_n|^p$, then it holds (3.13) because of the assumptions on g. Thus (3.14) is satisfied if (a) holds. Then, by Theorem 3.1, $f(x) - \bar{\mu}\dfrac{p}{q}(g(x) - 1) \geq \bar{\mu}$ for all $x \in C$, where $\bar{\mu} \doteq \min\{\|x\|^p :\ g(x) = 1, x \in C\} > 0$. By using a simple argument on the preceding inequality, one infers that $p \geq q$.

(b) $\Rightarrow$ (a): From (3.13) it follows that $\mu_0 \doteq \min\{f(x) :\ g(x) - 1 = 0,\ x \in C\} > 0$, and since $p \geq q$, strong duality for these data holds by Theorem 3.1. This means that $f(x) - \mu_0 \dfrac{p}{q}(g(x) - 1) \geq \mu_0 > 0$ for all $x \in C$, proving the desired result. □

3.4 Characterizing optimality conditions

We are now interested in obtaining necessary and/or sufficient optimality conditions of order zero, one or two, for local or global optimality for problem (3.10). All the results established in this section are new. In particular, in case f is a quadratic form and g the square of the Euclidean norm, Corollary 3.3 below enhances Proposition 3 of [130] since the convexity on C is not required. We point out that for quadratic optimization problems on a polyhedron, some optimality conditions were established in [24]. We refer to [18] for a method locating some particular local minima; some copositivity-based escape procedures for the StQO problem on the simplex are analyzed in [21].

3.4.1 Zero-order optimality conditions

We now provide a relationship between the minima of the original objective function and those of the Lagrangian, under strong duality and free of derivative.
Recall that $L(\lambda, x) \doteq f(x) + \lambda(g(x) - 1)$.

Theorem 3.2. *Let C be a closed cone, and f, g satisfy Assumption* (A)*. Then,*

(*a*) *Strong duality holds and* $\bar{x} \in \operatorname*{argmin}_K f$

$$\iff \bar{x} \in K \quad \textit{and} \quad \bar{x} \in \operatorname*{argmin}_C L\left(-\frac{p}{q} f(\bar{x}), \cdot\right).$$

(*b*) *Strong duality holds and* $0 \in \operatorname*{argmin}_C L\left(-\frac{p}{q}\mu_0, \cdot\right) \iff \mu_0(p - q) = 0.$

Proof. (a) $\Rightarrow$: This follows from (3.12):

$$L(-\frac{p}{q}\mu_0, \bar{x}) = f(\bar{x}) = \inf_{x \in C} \; L(-\frac{p}{q}\mu_0, x). \tag{3.15}$$

$\Leftarrow$: We have

$$\begin{aligned} \mu_0 &\leq f(\bar{x}) = L(-\frac{p}{q} f(\bar{x}), \bar{x}) = \inf_{x \in C} L(-\frac{p}{q} f(\bar{x}), x) \\ &\leq \inf_{x \in K} L(-\frac{p}{q} f(\bar{x}), x) = \inf_{x \in K} f(x) = \mu_0. \end{aligned}$$

Thus the proof of (a) is complete once one notices that:

$$L(-\frac{p}{q}\mu_0, \bar{x}) = \sup_{\lambda \in \mathbb{R}} \inf_{x \in C} \; L(\lambda, x).$$

(b): It is a consequence of (3.12):

$$\mu_0 = f(0) - \mu_0 \frac{p}{q}(g(0) - 1) = \mu_0 \frac{p}{q}.$$

□

When the assumption on strong duality is more precise, the following result is obtained.

Corollary 3.2. *Let C be a closed cone, and f, g satisfy Assumption* (A).

(*a*) *Assume that $\mu_0(p-q) > 0$. Then:*

$$\bar{x} \in \underset{C}{\operatorname{argmin}}\ L(-\frac{p}{q}\mu_0, \cdot) \Longleftrightarrow \bar{x} \in \underset{K}{\operatorname{argmin}}\ f.$$

(*b*) *Assume that $p = q > 0$. Then,*

$$0 \neq \bar{x} \in \underset{C}{\operatorname{argmin}}\ L(-\mu_0, \cdot) \implies \frac{1}{(g(\bar{x}))^{1/p}}\bar{x} \in \underset{K}{\operatorname{argmin}}\ f.$$

Proof. (*a*): By assumption on p, q, μ_0, strong duality holds. Thus, we need only to prove that $\bar{x} \in K$, and then the result follows from the previous theorem. We consider the case $\mu_0 > 0$ and $p > q > 0$ since the other is entirely similar. Suppose that $\bar{x} \notin K$; then $\bar{x} \neq 0$ due to the previous theorem, and so we can write set $\bar{x} = t\bar{y}$ for some $t > 0$ and $\bar{y} \in K$. By (3.15), we get:

$$\begin{aligned} \mu_0 &= \min_{x \in C} L(-\frac{p}{q}\mu_0, x) = f(\bar{x}) - \frac{p}{q}\mu_0(g(\bar{x}) - 1)) = t^p f(\bar{y}) - \frac{p}{q}\mu_0(t^q g(\bar{y}) - 1) \\ &\geq t^p\mu_0 - \frac{p}{q}\mu_0(t^q - 1). \end{aligned}$$

By defining $\varphi(\xi) \doteq \dfrac{\xi^q}{q} - \dfrac{\xi^p}{p}$, we obtain:

$$\varphi(t) = \frac{t^q}{q} - \frac{t^p}{p} \geq \varphi(1) > 0. \tag{3.16}$$

On the other hand, we have $\varphi'(\xi) < 0$ for all $\xi > 1$; $\varphi'(\xi) > 0$ for all $\xi \in]0, 1[$. Simple arguments show that the only possible value for t satisfying (3.16) is $t = 1$, which finally yields $\bar{x} \in K$.
(*b*): Simply use (3.12). □

3.4.2 KKT-points, L(agrange)-eigenvalues and second-order optimality conditions

We now derive first and second-order sufficient and/or necessary conditions for local or global optimality.
As usual, the notion of contingent cone will be needed. Given a set $M \subseteq \mathbb{R}^n$ and $x \in M$, the contingent cone of M at x, denoted by $T(M; x)$, is the set of vectors $v \in \mathbb{R}^n$ such that there exist $t_k > 0$, $x_k \in M$, $x_k \to x$, satisfying $t_k(x_k - x) \to v$. For a great account of its properties, we refer the book [11]. In general, we obtain:

$$T(M; x) \subseteq \overline{\text{cone}}(M - x),\ x \in M.$$

The equality is satisfied whenever M is convex.
It is known that the notion of KKT-point plays a crucial role for optimality. Thus, we assume that the functions f and g are differentiable at the reference point $\bar{x} \in C$.
Following [28] for instance, a point $x \in C$, $x \neq 0$, is said to be a KKT-point for problem (3.10) if there exists (Lagrangian multiplier) $\lambda \in \mathbb{R}$ such that:

$$\nabla f(x) - \lambda \frac{p}{q} \nabla g(x) \in (T(C;x))^*. \tag{3.17}$$

Let us denote the set of such λ associated to x by $\mathcal{L}(x)$. Actually $\mathcal{L}(x) = \left\{\frac{f(x)}{g(x)}\right\}$ whenever x is a KKT-point because of $\pm x \in T(C;x)$ and Euler identity.
In what follows, a collection of useful results is stated.

Remark 3.3. *Let f, g, C be as before. Given any $x \in C$, $x \neq 0$, the following assertions hold:*

(*i*) $\pm x \in T(C;x)$, *and so*

$$\nabla f(x) - \lambda \frac{p}{q} \nabla g(x) \in (T(C;x))^* \Longrightarrow f(x) = \lambda g(x);$$

(*ii*) *by Euler's identity,*

$$f(x) - \lambda g(x) = 0 \Longleftrightarrow x^\top \Big(\nabla f(x) - \lambda \frac{p}{q} \nabla g(x) \Big) = 0;$$

(*iii*) *if, in addition, C is convex,*

$$\nabla f(x) - \lambda \frac{p}{q} \nabla g(x) \in (T(C;x))^* \Longleftrightarrow \begin{cases} \nabla f(x) - \lambda \frac{p}{q} \nabla g(x) \in C^*; \\ f(x) = \lambda g(x). \end{cases}$$

Motivated by the previous remark, some notions are introduced:

Definition 3.1. *We say that $\lambda \in \mathbb{R}$ is (f,g)-eigenvalue (or simply, eigenvalue) if there exists $x \in C$, $x \neq 0$, such that $f(x) = \lambda g(x)$. The dependence of f and g will be omitted when no confusion arises. We say that $\lambda \in \mathbb{R}$ is L(agrange)-eigenvalue if there exists $x \in C$, $x \neq 0$ such that $\lambda \in \mathcal{L}(x)$. The set of those x is denoted by $\mathcal{K}(\lambda)$, and so, λ is L-eigenvalue if and only if $\mathcal{K}(\lambda) \neq \emptyset$. Every x in $\mathcal{K}(\lambda)$ is called a L-eigenvector associated to λ. A pair $(\lambda, x) \in \mathbb{R} \times (C \setminus \{0\})$ with $x \in \mathcal{K}(\lambda)$ (or, equivalently, $\lambda \in \mathcal{L}(x)$) is called L-eigenpair.*

In view of Remark 3.3, under convexity on C, the L-eigenvalue problem reduces to the homogeneous complementarity problem:

$$\begin{aligned} \nabla f(x) - \lambda \frac{p}{q} \nabla g(x) &\in C^* \\ (\nabla f(x) - \lambda \frac{p}{q} \nabla g(x))^\top x &= 0, \\ x \in C,\ x &\neq 0. \end{aligned}$$

From which we infer:

$$P_C(x - \nabla f(x) + \lambda \frac{p}{q} \nabla g(x)) = x; \quad P_{C^0}(x - \nabla f(x) + \lambda \frac{p}{q} \nabla g(x)) = -\nabla f(x) + \lambda \frac{p}{q} \nabla g(x),$$

where $C^0 = -C^*$, $P_M(v)$ stands for the orthogonal projection of v onto M. This is certainly the basis for some proximal-point algorithms.

The next remark lists some basic facts from standard convex analysis.

Remark 3.4. *Let $\emptyset \neq C \subseteq \mathbb{R}^n$ be a cone. We obtain:*

(*i*) $C - C \subseteq \text{span } C = \text{aff } C$;

(*ii*) *if C is convex then $\overline{\text{cone}}(C - \bar{x}) = \overline{C + \mathbb{R}\bar{x}}$, and therefore $[\overline{\text{cone}}(C - \bar{x})]^* = C^* \cap \bar{x}^\perp$;*

(*iii*) *if $\bar{x} \in \text{ri } C$ then $\text{cone}(C - \bar{x}) = \text{span } C = C - \bar{x}$, and so $[\text{cone}(C - \bar{x})]^* = (\text{span } C)^\perp = (C - \bar{x})^\perp = C^\perp$.*

We start with the following new second-order necessary optimality condition for local optimality. This result asserts that every local optimal solution to problem (3.10) is a KKT-point provided either C is convex or $\bar{x} \in \text{ri } C$, and so, $f(\bar{x})$ is L-eigenvalue. Theorem 1 in [130] is a special case of our result, when, besides the convexity of C, f is a quadratic form and g is the square of the Euclidean norm.

Theorem 3.3. *Let C be a closed cone, and f, g satisfy Assumption* (A) *with both functions being twice differentiable at $\bar{x}$, where $\bar{x}$ is any local solution to problem* (3.10)*. The following statements hold.*

(*a*) *If either C is convex or $\bar{x} \in \text{ri } C$, then*

$$\nabla f(\bar{x}) - \frac{p}{q} f(\bar{x}) \nabla g(\bar{x}) \in [\overline{\text{cone}}(C - \bar{x})]^*.$$

As a consequence, $(f(\bar{x}), \bar{x})$ is an L-eigenpair. In other words, $\bar{x}$ is a KKT-point having $f(\bar{x})$ as Lagrange multiplier;

(*b*) *if C is convex, the matrix:*

$$\nabla^2 f(\bar{x}) - \frac{p}{q} f(\bar{x}) \nabla^2 g(\bar{x}) - \frac{p}{q}\Big(\frac{p}{q} - 1\Big) f(\bar{x}) \nabla g(\bar{x}) \nabla g(\bar{x})^\top \text{ is copositive on } \mathcal{D}(\bar{x}),$$

where $\mathcal{D}(\bar{x}) \doteq \overline{[\nabla f(\bar{x}) - \frac{p}{q} f(\bar{x}) \nabla g(\bar{x})]^\perp \cap (\text{cone}(C - \bar{x}))}$.

(*c*) *If $\bar{x} \in \text{ri } C$, then*

$$\nabla^2 f(\bar{x}) - \frac{p}{q} f(\bar{x}) \nabla^2 g(\bar{x}) - \frac{p}{q}\Big(\frac{p}{q} - 1\Big) f(\bar{x}) \nabla g(\bar{x}) \nabla g(\bar{x})^\top \text{ is copositive on } \overline{\text{cone}}(C - \bar{x}).$$

Proof. Let $\bar{x}$ be a local solution to problem (3.10), that is, $f(\bar{x}) \leq f(x)$ for all $x \in C \cap U_0$ satisfying $g(x) = 1$, for some open neighborhood, U_0, of $\bar{x}$.
(*a*): It is known that in case C is convex, given any $v \in C - \bar{x}$, we can choose $\varepsilon \in]0, 1[$ such that:

$$\frac{\bar{x} + tv}{(g(\bar{x} + tv))^{1/q}} \in C \cap U_0, \quad \forall\, t \in]0, \varepsilon[.$$

Set

$$\phi(t) \doteq f\Big(\frac{\bar{x}+tv}{(g(\bar{x}+tv))^{1/q}}\Big) = \frac{1}{(g(\bar{x}+tv))^{p/q}} f(\bar{x}+tv).$$

By assumption, $\phi(0) \leq \phi(t)$ for all $t \in]0, \varepsilon[$. This implies that the right-derivative of ϕ at 0 is nonnegative:

$$0 \leq \phi'(0) = \Big(\nabla f(\bar{x}) - \frac{p}{q} f(\bar{x}) \nabla g(\bar{x})\Big)^{\top} v.$$

It is valid for every $v \in C - \bar{x}$, and so the conclusion follows provided C is convex.
We now consider the case when $\bar{x} \in \text{ri } C$. This means that $U_0 \cap \text{aff } C = U_0 \cap (C - C) \subseteq C$ for some open neighborhood, U_0, of $\bar{x}$. From this, we derive the same result as in the convex case for the function:

$$\phi(t) \doteq f\Big(\frac{\bar{x}-tv}{(g(\bar{x}-tv))^{1/q}}\Big),$$

since $\text{cone}(C - \bar{x})$ is a subspace.
(b): Let $v \in [\nabla f(\bar{x}) - \frac{p}{q} f(\bar{x}) \nabla g(\bar{x})]^{\perp} \cap (C - \bar{x})$. We use the Maclaurin expansion for the function ϕ:

$$\phi(t) = \phi(0) + t\phi'(0) + \frac{1}{2}\phi''(0)t^2 + t^2 o(t),$$

where $o(t) \to 0$ as $t \to 0^+$. Then:

$$0 \leq \phi''(0) = v^{\top}\Big(\nabla^2 f(\bar{x}) - \frac{p}{q} f(\bar{x}) \nabla^2 g(\bar{x}) - \frac{p}{q}\Big(\frac{p}{q} - 1\Big) f(\bar{x}) \nabla g(\bar{x}) \nabla g(\bar{x})^{\top}\Big) v.$$

From which, the desired result is obtained.
(c): Simply notice that the function ϕ introduced in (a) is defined in $]-\varepsilon, \varepsilon[$, and by local opimality, $\phi'(0) = 0$. □

A remark must be emphasized.

Remark 3.5. *As was pointed out in Theorem 3.3, every local solution, $\bar{x}$, is not necessarily a KKT-point, even being global, except in the cases when either C is convex or $\bar{x} \in \text{ri } C$. Indeed, Example 3.2 shows an instance where every global solution is not a KKT-point. Here C is nonconvex and certainly the global solution is not in the relative interior of C. Every solution will be a KKT-point under strong duality, as Theorem 3.6 below shows. When no extra-conditions are imposed on the data, a geometric characterization of KKT-points is established in (b) of Theorem 3.6.*

Results of the following kind are standard, but you can see how the structure of our problem takes place.
Recall that $L(\lambda, x) \doteq f(x) + \lambda(g(x) - 1)$ and $\nabla_x L(\lambda, x) \doteq \nabla f(x) + \lambda \nabla g(x)$.

Theorem 3.4. *Let C be a closed cone, and f, g satisfy Assumption* (A) *with both functions being twice differentiable at $\bar{x} \in K$. If,*

$$\nabla_x L(-\frac{p}{q} f(\bar{x}), \bar{x}) \in [\overline{\text{cone}}(C - \bar{x})]^*,$$

and the matrix:

$$\nabla_x^2 L(-\frac{p}{q} f(\bar{x}), \bar{x}) \quad \textit{is strictly copositive on } \overline{\text{cone}}(C - \bar{x}),$$

then $\bar{x}$ is a strict local minimum of $L(-\frac{p}{q} f(\bar{x}), \cdot)$ on C.

Theorem 3.5. *Let C be a convex closed cone, and f, g satisfy Assumption* (A) *with both functions being twice differentiable at $\bar{x} \in K$. Assume that*

$$\nabla_x L(-\frac{p}{q} f(\bar{x}), \bar{x}) \in [\overline{\text{cone}}(C - \bar{x})]^*.$$

(a) *If $L(-\frac{p}{q} f(\bar{x}), \cdot)$ is pseudoconvex then $\bar{x}$ is a solution to problem* (3.10)*;*

(b) *If f is pseudoconvex and the function $x \mapsto -\frac{p}{q} f(\bar{x}) g(x)$ is quasiconvex, then $\bar{x}$ is a solution to problem* (3.10)*.*

An important consequence of Theorem 3.3 concerns the particular case when f and g are quadratic forms. The next corollary does not require convexity on C as Proposition 3 in [130] does.

Corollary 3.3. *Let C be a closed cone, and $f(x) = x^\top A x$, $g(x) = x^\top x$ with $A = A^\top$. Assume that $\bar{x} \in \text{ri } C$ is a local solution to problem* (3.10)*. Then $\bar{x}$ is a global solution.*

Proof. For every $x \in C$, we have

$$\begin{aligned} L(-f(\bar{x}), x) - L(-f(\bar{x}), \bar{x}) &= \nabla_x L(-f(\bar{x}), \bar{x})^\top (x - \bar{x}) + \frac{1}{2}(x - \bar{x})^\top \nabla_x^2 L(-f(\bar{x}), \bar{x})(x - \bar{x}) \\ &= \frac{1}{2}(x - \bar{x})^\top \nabla_x^2 L(-f(\bar{x}), \bar{x})(x - \bar{x}). \end{aligned}$$

We now apply Theorem 3.3 together with the above expansion to conclude that:

$$\bar{x} \in \operatorname*{argmin}_C L(-f(\bar{x}), \cdot).$$

The result is obtained as a consequence of (a) in Theorem 3.2. □

In what follows we consider the case $C = \mathbb{R}^n_+$.
Set $I \doteq \{1, \ldots, n\}$. Given $J \subseteq I$, any $x \in \mathbb{R}^n$ is written as $x = (x_J, x_{-J})$ where $x_J \doteq (x_i)_{i \in J}$, that is, x_J is the vector with components whose indexes belong to J; and x_{-J} is the vector with the remaining components. Thus $0_{-J} \in \mathbb{R}^{n-|J|}$ is the vector with all its components being zero and indexes in $I \setminus J$, where $|J|$ means the cardinality of J.
The next result is expected.

Proposition 3.3. *Let f, g be differentiable functions satisfying Assumption* (A) *with $C = \mathbb{R}^n_+$, and $\lambda \in \mathbb{R}$. Then:*

$$\lambda \text{ is } L-\text{eigenvalue} \iff \begin{cases} \exists\ \emptyset \neq J \subseteq I,\ \exists\ \bar{y} \in \mathbb{R}^{|J|}_{++}: \\ \nabla_J f(\bar{y}, 0_{-J}) - \lambda \dfrac{p}{q} \nabla_J g(\bar{y}, 0_{-J}) = 0; \\ \nabla_{-J} f(\bar{y}, 0_{-J}) - \lambda \dfrac{p}{q} \nabla_{-J} g(\bar{y}, 0_{-J}) \in \mathbb{R}^{n-|J|}_+. \end{cases}$$

Here, $\nabla_J f(x)$ (resp. $\nabla_{-J} f(x)$) stands for the vector whose components are the partial derivatives of f at $x \in \mathbb{R}^n$ with respect to the indexes belonging to J (resp. $I \setminus J$).

Proof. $\Rightarrow$: Let $\bar{x} \in \mathcal{K}(\lambda)$. Then $\bar{x} \in \mathbb{R}^n_+ \setminus \{0\}$, and set $J \doteq \{i \in I :\ \bar{x}_i > 0\}$. Thus $\bar{x} = (\bar{x}_J, 0_{-J})$ and therefore:

$$\nabla f(\bar{x}) - \lambda \frac{p}{q} \nabla g(\bar{x}) \in [T(\mathbb{R}^n_+; \bar{x})]^* = \{0_J\} \times \mathbb{R}^{n-|J|}_+,$$

since:

$$T(\mathbb{R}^n_+; \bar{x}) = T(\mathbb{R}^{|J|}_+; \bar{x}_J) \times T(\mathbb{R}^{n-|J|}_+; 0_{-J}).$$

By taking $\bar{y} = \bar{x}_J \in \mathbb{R}^{|J|}_{++}$, the desired implication is proved.
$\Leftarrow$: By setting $\bar{x} = (\bar{y}_J, 0_{-J})$, we obtain:

$$\nabla f(\bar{x}) - \lambda \frac{p}{q} \nabla g(\bar{x}) \in \{0_J\} \times \mathbb{R}^{n-|J|}_+ = [T(\mathbb{R}^n_+; \bar{x})]^*.$$

This means that λ is L-eigenvalue. □

Since ∇h is positively homogeneous with degree $p-1$, the next assertions are easy to check.

- $x \in \mathcal{K}(\lambda) \iff \lambda \in \mathcal{L}(x)$;
- $x \in K$, $\mu_0 \in \mathcal{L}(x) \Longrightarrow x \in \operatorname*{argmin}_K f \Longrightarrow \mathcal{L}(x) \subseteq \{\mu_0\}$;
- given any $x \in C$, $x \neq 0$, it holds ($p \neq q$):

$$T(C; x) = T(C; tx)\ \ \forall\ t > 0;\ \ \lambda \in \mathcal{L}(x) \iff \lambda t^{p-q} \in \mathcal{L}(tx)\ \ \forall\ t > 0. \tag{3.18}$$

The next result establishes a relationship between both types of eigenvalues under strong duality.

Proposition 3.4. *Let C be a closed cone, and f, g be differentiable functions satisfying Assumption* (A). *Then,*

(*a*) $\mu_0 = \min\{\lambda \in \mathbb{R} :\ \lambda \text{ is eigenvalue}\}$;

(*b*) $\mu_0 = \min\{\lambda \in \mathbb{R} :\ \lambda \text{ is } L-\text{eigenvalue}\}$, *provided strong duality holds* (*see Theorem 3.1*).

Proof. (a): It is straightforward.
(b): By (a), we only need to check the inequality "$\geq$". Take any $\bar{x} \in \operatorname*{argmin}_K f$. The usual necessary optimality condition along with Theorem 3.2 allow us to infer that $\nabla f(\bar{x}) - \mu_0 \frac{p}{q} \nabla g(\bar{x}) \in [T(C;\bar{x})]^*$, which shows that μ_0 is L-eigenvalue, and so the proof is complete. □

The next remark refers to eigenvalues analysis of symmetric (supersymmetric) tensors.

Remark 3.6. *One important implication from Proposition 3.4 concerns the existence of L-eigenvalues of a real m-order n-dimensional symmetric tensor. See Sect. 3.7 for details.*

Notice that, in principle, a KKT-point is defined also for infeasible points. In order to characterize the validity of the KKT optimality conditions, (3.17), for problem (3.10), we need to consider the linearized approximation problem defined, given $\bar{x} \in C$, by:

$$\mu_L \doteq \inf_{v \in G_0(\bar{x})} \nabla f(\bar{x})^\top v, \tag{3.19}$$

where:

$$G_0(\bar{x}) \doteq \Big\{ v \in T(C;\bar{x}) : \ \nabla g(\bar{x})^\top v = 0 \Big\}.$$

In our model, we have $\nabla g(x) \neq 0$ for all $x \in C$, $x \neq 0$ because of the Euler identity.

Set $F_L(v) \doteq (\nabla g(\bar{x})^\top v, \nabla f(\bar{x})^\top v)$. It is obvious that $\mu_L \in \{-\infty, 0\}$, and:

$$\begin{aligned} \mu_L = 0 &\iff [v \in T(C;\bar{x}),\ \nabla f(\bar{x})^\top v < 0 \Longrightarrow \nabla g(\bar{x})^\top v \neq 0] && (3.20)\\ &\iff F_L(T(C;\bar{x})) \cap -(\{0\} \times \mathbb{R}_{++}) = \emptyset \\ &\iff [F_L(T(C;\bar{x})) + \mathbb{R}_+(0,1)] \cap (-\{0\} \times \mathbb{R}_{++}) = \emptyset. && (3.21) \end{aligned}$$

Some important facts on the set $F_L(T(C;\bar{x})) + \mathbb{R}_+(0,1)$ are collected in the next remark.

Remark 3.7. *With the above data and notation, we obtain the following*

(i) $\lambda \in \mathcal{L}(\bar{x}) \iff (-\lambda \frac{p}{q}, 1) \in [F_L(T(C;\bar{x})) + \mathbb{R}_+(0,1)]^* \iff (\lambda, \bar{x})$ *is L-eigenpair.*

(ii) *Assume that $F_L(T(C;\bar{x})) + \mathbb{R}_+(0,1)$ is convex. Then, either*

1. $F_L(T(C;\bar{x})) + \mathbb{R}_+(0,1) = \mathbb{R}^2$, *or*
2. $F_L(T(C;\bar{x})) + \mathbb{R}_+(0,1) = \Big\{ (v,w) : \ w \geq \frac{p}{q} \frac{f(\bar{x})}{g(\bar{x})} v \Big\}$.

   Indeed, since $\pm\bar{x} \in T(C;\bar{x})$, $F_L(\pm\bar{x}) = \pm(\nabla g(\bar{x})^\top \bar{x}, \nabla f(\bar{x})^\top \bar{x})$, and so, $\mathbb{R}\Big(1, \frac{p}{q} \frac{f(\bar{x})}{g(\bar{x})}\Big) \subseteq F_L(T(C;\bar{x}))$. This implies

$$\mathbb{R}\Big(1, \frac{p}{q} \frac{f(\bar{x})}{g(\bar{x})}\Big) + \mathbb{R}_+(0,1) \subseteq F_L(T(C;\bar{x})) + \mathbb{R}_+(0,1).$$

From which the conclusion follows.

We are now ready to describe, firstly, a new necessary and sufficient condition for a non-zero vector to be a KKT-point; this shows, looking at carefully its proof, that a Fritz John point is indeed a KKT-point, thanks to the structure of our model. Secondly, under strong duality, it establishes that every minimizer is a KKT-point. Thus, next theorem supplements the results stated in Theorem 3.3.

Theorem 3.6. *Let C be a closed cone, and f, g be differentiable functions satisfying Assumption* (A)*. Assume that $\bar{x} \in C$, $\bar{x} \neq 0$. The following assertions hold:*

(*a*) *$\bar{x}$ is a KKT-point $\Longleftrightarrow$ $F_L(T(C;\bar{x})) + \mathbb{R}_+(0,1)$ is different from $\mathbb{R}^2$ and convex.*

In such a case:

$$F_L(T(C;\bar{x})) + \mathbb{R}_+(0,1) = \Big\{(v,w) : \ w \geq \frac{p}{q}\frac{f(\bar{x})}{g(\bar{x})}v\Big\}.$$

(*b*) *if strong duality holds, then, for $\bar{x} \in K$,*

$$\bar{x} \in \operatorname*{argmin}_K f \Longleftrightarrow \mu_0 \in \mathcal{L}(\bar{x})(= \{\mu_0\}) \Longleftrightarrow \bar{x} \in \mathcal{K}(\mu_0).$$

Proof. (*a*) $\Rightarrow$: $\bar{x}$ is a KKT-point if and only if there exists $\lambda \in \mathbb{R}$ such that:

$$\inf_{v \in T(C;\bar{x})} \Big\langle \nabla f(\bar{x}) - \lambda \frac{p}{q}\nabla g(\bar{x}), v \Big\rangle \geq 0 \geq \inf_{v \in G_0(\bar{x})} \nabla f(\bar{x})^\top v.$$

Thus, $\bar{x} \in$ is a KKT-point if and only if $\mu_L = 0$ and (strong duality for (3.19) holds):

$$\inf_{v \in T(C;\bar{x})} \Big\langle \nabla f(\bar{x}) - \lambda \frac{p}{q}\nabla g(\bar{x}), v \Big\rangle = \inf_{v \in G_0(\bar{x})} \nabla f(\bar{x})^\top v = \mu_L.$$

We have already noticed that $\mu_L = 0$ is equivalent to (3.20); which means that $F_L(T(C;\bar{x})) + \mathbb{R}_+(0,1) \neq \mathbb{R}^2$. The convexity of $F_L(T(C;\bar{x})) + \mathbb{R}_+(0,1)$ follows from Proposition 5.1 in [58], when applying to $A = F_L(T(C;\bar{x}))$ and $P = \mathbb{R}_+(0,1)$.
$\Leftarrow$: By Remark 3.7,

$$\mathbb{R}\Big(1, \frac{p}{q}\frac{f(\bar{x})}{g(\bar{x})}\Big) + \mathbb{R}_+(0,1) = F_L(T(C;\bar{x})) + \mathbb{R}_+(0,1),$$

which yields (3.21). By applying a separation result on convex sets to (3.21), we have the existence of $\alpha \in \mathbb{R}$, $\beta \geq 0$, $(\alpha,\beta) \neq (0,0)$ such that $\beta\nabla f(\bar{x}) + \alpha\nabla g(\bar{x}) \in [T(C;\bar{x})]^*$. By Remark 3.3, $\beta p f(\bar{x}) + \alpha q g(\bar{x}) = 0$. Thus, if $\beta = 0$ then $\alpha = 0$ since $g(\bar{x}) \neq 0$, and so $\beta > 0$. Hence, the result is obtained.
(*b*): It follows from Theorem 3.2 and the standard first-order necessary optimality condition for $x \in \operatorname*{argmin}_C L\Big(-\mu_0\frac{p}{q}, \cdot\Big)$. □

The next example shows an application of (a) in Theorem 3.6 without having strong duality, and so (b) of the same theorem is not applicable; whereas Example 3.2 shows the result in (b) may be false if strong duality fails.

Example 3.1. (*Strong duality fails, and so* (b) *of Theorem 3.6 is not applicable, but* (a) *so is*) *Let* α *and* β *be any positive real numbers. We will analyze the problem:*

$$\mu_0(\alpha,\beta) \doteq \min\Big\{f(x):\ g_{\alpha,\beta}(x)=1,\ x\in\mathbb{R}^2\Big\},$$

with the data $f(x_1,x_2)=x_1+2x_2$ *and* $g_{\alpha,\beta}(x)=\sqrt[4]{\alpha x_1^2+\beta x_2^2}$ *and* $C=\mathbb{R}^2$. *In this case,* $p=1$, $q=1/2$ *and* $T(C;x)=\mathbb{R}^2$ *for all* $x\in\mathbb{R}^2$. *By virtue of* (3.18), *we search KKT-points only in the feasible set* K. *Thus, from* (3.17), *we get* $1-\lambda\alpha x_1=0=2-\lambda\beta x_2$, *implying* $\lambda\neq 0$, $x_1\neq 0\neq x_2$, *and therefore* $x_2=\dfrac{2\alpha}{\beta}x_1$. *For such points* x *and* $(v_1,v_2)\in\mathbb{R}^2$, *we obtain:*

$$F_L(v)=\Big(\frac{\alpha}{2}x_1v_2+\frac{\beta}{2}x_2v_2,v_1+2v_2\Big)=\Big(\frac{\alpha}{2}x_1,1\Big)\Big(v_1+2v_2\Big).$$

Then $F_L(T(C;x))=\Big\{(w_1,w_2):\ w_2=\dfrac{2}{\alpha x_1}w_1,w_1\in\mathbb{R}\Big\}$, *implying the convexity of* $F_L(T(C;x))+\mathbb{R}_+(0,1)$ *and* (3.21). *Hence, by Theorem 3.6,* $(x_1,x_2)=x_1\Big(1,\dfrac{2\alpha}{\beta}\Big)$ *is a KKT-point with* $g_{\alpha,\beta}(x_1,x_2)=1$. *That is,*

$$\Big\{x\in K:\ \mathcal{L}(x)\neq\emptyset\Big\}=\left\{\pm\sqrt{\frac{\beta}{\alpha\beta+4\alpha^2}}\left(1,\frac{2\alpha}{\beta}\right)\right\}.$$

It is easy to check that one is the minimizer (and the other the maximizer) with minimum value:

$$\mu_0=\mu_0(\alpha,\beta)=-\sqrt{\frac{\beta+4\alpha}{\alpha\beta}}.$$

By applying the second-order optimality condition from Theorem 3.3, we discard those points which are not local minima. For this example, we get:

$$\nabla^2 f(\bar{x})-\frac{p}{q}f(\bar{x})\nabla^2 g(\bar{x})-\frac{p}{q}\Big(\frac{p}{q}-1\Big)f(\bar{x})\nabla g(\bar{x})\nabla g(\bar{x})^\top$$

$$=\ -2f(\bar{x})\Big(\nabla^2 g(\bar{x})+\nabla g(\bar{x})\nabla g(\bar{x})^\top\Big)=-2f(\bar{x})\begin{pmatrix}\dfrac{2\alpha^2}{\beta+4\alpha} & -\dfrac{\alpha\beta}{\beta+4\alpha}\\ -\dfrac{\alpha\beta}{\beta+4\alpha} & \dfrac{\beta^2}{2(\beta+4\alpha)}\end{pmatrix}.$$

This is positive semidefinite if and only if $f(\bar{x})\leq 0$, *since the matrix*

$$\begin{pmatrix}\dfrac{2\alpha^2}{\beta+4\alpha} & -\dfrac{\alpha\beta}{\beta+4\alpha}\\ -\dfrac{\alpha\beta}{\beta+4\alpha} & \dfrac{\beta^2}{2(\beta+4\alpha)}\end{pmatrix}$$

is positive semidefinite, having as eigenvalues 0 *and* $\dfrac{4\alpha^2+\beta^2}{2(\beta+4\alpha)}$. *This second-order condition is satisfied only for the point:*

$$\bar{x} = -\sqrt{\frac{\beta}{\alpha\beta+4\alpha^2}}\left(1, \frac{2\alpha}{\beta}\right).$$

Finally, by using Theorem 3.5, we conclude that such a point is, in fact, a minimizer.

The next example is also related with Theorem 3.3.

Example 3.2. (*Strong duality fails,* C *is not convex, every minimizer is not a KKT-point, and* $\operatorname{ri} C = \emptyset$)
Take the functions:

$$f(x_1,x_2,x_3) \doteq \frac{(x_1^2+x_2^2+\frac{1}{4}x_3^2)^4}{\sqrt{x_1+x_3}} - \sqrt{x_3^{15}},\ g(x_1,x_2,x_3) \doteq \sqrt[8]{x_1^2+x_2^2+\frac{1}{4}x_3^2}.$$

and the closed, nonconvex, cone:

$$C \doteq \left\{(x_1,x_2,x_3)\in\mathbb{R}^3 : x_1\geq 0, x_2\geq 0, x_3\geq 0,\ x_1x_2=0\right\}.$$

Clearly, Assumption (A) fulfills for these data, with $p=15/2$, $q=1/4$. *For* $\bar{x}\in K$, *we obtain:*

$$f(\bar{x}_1,\bar{x}_2,\bar{x}_3) = \frac{1}{\sqrt{\bar{x}_1+\bar{x}_3}} - \sqrt{\bar{x}_3^{15}}, \operatorname*{argmin}_K f = \{(0,0,2)\}, \mu_0 = f(\bar{x}) = \frac{1}{\sqrt{2}} - \sqrt{2^{15}} < 0.$$

Therefore strong duality does not hold. Moreover,

$$T(C,\bar{x}) \doteq \{(v_1,v_2,v_3)\in\mathbb{R}^3 :\ v_1v_2=0, v_1\geq 0, v_2\geq 0, v_3\in\mathbb{R}\}.$$
$$[T(C,\bar{x})]^* \doteq \{(v_1,v_2,v_3)\in\mathbb{R}^3 :\ v_1\geq 0, v_2\geq 0, v_3=0\}.$$

Since $f_{x_1}(\bar{x}) = -\dfrac{1}{2\bar{x}_3^{3/2}} < 0$ *and* $g_{x_1}(\bar{x}) = 0$, *we infer that there is no* $\lambda\in\mathbb{R}$ *such that* $\nabla f(\bar{x}) - \lambda\dfrac{p}{q}\nabla g(\bar{x}) \in [T(C,\bar{x})]^*$. *We also obtain:*

$$F_L(v) \doteq \begin{pmatrix}\nabla g(\bar{x})^\top v\\ \nabla f(\bar{x})^\top v\end{pmatrix} = \begin{pmatrix} v_3 g_{x_3}(\bar{x})\\ v_1 f_{x_1}(\bar{x}) + v_3 f_{x_3}(\bar{x})\end{pmatrix}.$$

Then $F_L(T(C;\bar{x})) = \left\{(u_1,u_2)\in\mathbb{R}^2 : u_2 \leq \dfrac{f_{x_3}(\bar{x})}{g_{x_3}(\bar{x})}u_1, u_1\in\mathbb{R}\right\}$, *where* $g_{x_3}(\bar{x}) = 1/8$ *and* $f_{x_3}(\bar{x}) = -676.17$, *which yields* $F_L(T(C;\bar{x})) + \mathbb{R}_+(0,1) = \mathbb{R}^2$. *This means that* (3.21), *or equivalently,* (3.20) *is not satisfied.*

Let us establish a necessary and sufficient condition for $\bar{x}$ to be a KKT-point when it is a minimizer of f on K. By virtue of Lemma 3.1, we split our discussion into two cases: $\mu_0 < 0$ and $\mu_0 > 0$. The remaining case $\mu_0 = 0$, which implies strong duality, can be dealt with (b) of Theorem 3.6. We need the following notation:

$$F_L^-(v) \doteq (-\nabla g(\bar{x})^\top v, \nabla f(\bar{x})^\top v),$$

and the condition:

$$[v_k \in T(C;\bar{x}), \nabla g(\bar{x})^\top v_k \to 0, \nabla f(\bar{x})^\top v_k < 0] \Longrightarrow \limsup_k \nabla f(\bar{x})^\top v_k = 0. \tag{3.22}$$

Notice that Example 3.1 also shows that condition (3.22) cannot be substituted by (3.20) in the following theorem.

Theorem 3.7. *Let C be a closed cone, and f, g be differentiable functions at $\bar{x} \in K$, satisfying Assumption* (A)*.*

(*a*) *Assume that $\mu_0 < 0$ and $\bar{x} \in \operatorname{argmin}_K f$. Then,*

$$\bar{x} \text{ is a } KKT-\text{point } \Longleftrightarrow F_L(T(C;\bar{x})) + \mathbb{R}^2_+ \text{ is convex and (3.22) is satisfied}].$$

(*b*) *Assume that $\mu_0 > 0$ and $\bar{x} \in \operatorname{argmin}_K f$. Then,*

$$\bar{x} \text{ is a } KKT-\text{point } \Longleftrightarrow F_L^-(T(C;\bar{x})) + \mathbb{R}^2_+ \text{ is convex and (3.22) is satisfied}].$$

Proof. See Corollary 5.5 in [57] and Lemma 3.1. □

A more verifiable condition than (3.22) is given next:

Proposition 3.5. *With the data as above, we have:*

$$(3.23) \Longrightarrow (3.22) \Longrightarrow \mu_L = 0,$$

where:

$$[v_k \in T(C;\bar{x}), v_k \to v \neq 0, \nabla f(\bar{x})^\top v_k < 0] \Longrightarrow \nabla g(\bar{x})^\top v \neq 0. \tag{3.23}$$

Proof. Firstly, we easily obtain that (3.22) implies $\mu_L = 0$ (simply take the constant sequence $v_k = v$ to get $\mu_L = 0$).
(3.23) $\Rightarrow$ (3.22): Let $v_k \in T(C;\bar{x})$ such that $\|v_k\| \to +\infty$, $\nabla g(\bar{x})^\top v_k \to 0$, $\nabla f(\bar{x})^\top v_k < 0$. In case $v_k \to 0$, we get (3.22) obviously holds. Now, two possibilities arise: $\sup_{k\in\mathbb{N}} \|v_k\| < +\infty$ with $v_k \not\to 0$ and $\sup_{k\in\mathbb{N}} \|v_k\| = +\infty$. Under the first possibility, up to a subsequence, we get $v_k \to v \neq 0$. In such a case, since $\nabla f(\bar{x})^\top v_k < 0$, by (3.23) we obtain $\nabla g(\bar{x})^\top v > 0$, yielding a contradiction.
If second possibility holds, up to a subsequence, we may suppose that $\dfrac{v_k}{\|v_k\|} \to v_0 \neq 0$. Since $\nabla f(\bar{x})^\top \dfrac{v_k}{\|v_k\|} < 0$ by (3.23) we get $\nabla g(\bar{x})^\top v_0 \neq 0$. Moreover, since $\nabla g(\bar{x})^\top v_k \to 0$, we have $\nabla g(\bar{x})^\top \dfrac{v_k}{\|v_k\|} \to 0 = \nabla g(\bar{x})^\top v_0$, which contradicts the fact that $\nabla g(\bar{x})^\top v_0 \neq 0$. This proves that under (3.23), the conditions in the left-hand side of (3.22) are fulfilled only if $v_k \to 0$, so that (3.22) holds. □

3.4.3 The case when C comes from two quadratic forms

In most applications the closed cone C is defined by some quadratic forms. We will consider the case of two forms:

$$C \doteq \{x \in \mathbb{R}^n : \ x^\top B_1 x = 0, \ x^\top B_2 x = 0\}.$$

Here B_i, $i = 1, 2$, are symmetric matrices of order n with real entries. Notice that any set of the form $\{x \in \mathbb{R}^n : \ x^\top B_i x \leq 0\}$ can be reduced to a set with equality, simply by adding one additional component:

$$\{(x, x_{n+1}) \in \mathbb{R}^{n+1} : \ x^\top B_i x + x_{n+1}^2 = 0\} = \{x \in \mathbb{R}^{n+1} : \ x^\top \tilde{B}_i x = 0\},$$

for a suitable $\tilde{B}_i$, as usual.
We start by a simple standard result:

Proposition 3.6. *Let $R \subseteq \mathbb{R}^n$ be any nonempty set and S be a vector subspace in $\mathbb{R}^n$. Then:*

$$(R \cap S)^* = R^* + S^\perp.$$

Proof. We only need to check one inclusion. Take any $z \in (R \cap S)^*$ and write $z = z_1 + z_2$ with $z_1 \in S$ and $z_2 \in S^\perp$. We claim that $z_1 \in R^*$. Let $\xi \in R$. If $\xi \in S$ then:

$$z_1^\top \xi = z^\top \xi - z_2^\top \xi = z^\top \xi \geq 0.$$

In case, $\xi \in S^\perp$, then $z_1^\top \xi = 0$. Thus, the claim is proved, and the proof is complete. □

Let $C_i \doteq \{x \in \mathbb{R}^n : \ x^\top B_i x = 0\}$, for $i = 1, 2$. Then $C_i = -C_i$, and so $\overline{\text{co}}\ C_i$ is a vector subspace. For a given $\bar{x} \in C_i$, it is well known that (see for instance [42]):

$$T(C_i; \bar{x}) = (B_i \bar{x})^\perp, \ \text{if } \bar{x} \notin \ker B_i; \ T(C_i; \bar{x}) = C_i = C_i - \bar{x}, \ \text{if } \bar{x} \in \ker B_i.$$

and, in case B_i is indefinite, we have [115, Lemma 3.10]:

$$\text{co}\ C_i = \mathbb{R}^n = \text{span}\ C_i.$$

Proposition 3.7. *Let C_i, $i = 1, 2$, be as above, and $\bar{x} \in C_1 \cap C_2$. Then $T(C_1 \cap C_2; \bar{x}) = T(C_1; \bar{x}) \cap T(C_2; \bar{x})$, under any of the following circumstances:*

(*a*) *$\{\nabla g_1(\bar{x}), \nabla g_2(\bar{x})\}$ is linearly independent.*

(*b*) *$\nabla g_1(\bar{x}) = 0 = \nabla g_2(\bar{x})$.*

Proof. (*a*): It is [42, Theorem 2.2].
(*b*): In this case, we have $C_i = C_i \pm \bar{x} = T(C_i; \bar{x})$ for $i = 1, 2$. Thus, if $v \in T(C_1; \bar{x}) \cap T(C_2; \bar{x})$, then $(\bar{x} + tv)^\top B_i (\bar{x} + tv) = 0$ for all $t > 0$ and $i = 1, 2$, which mean that $\bar{x} + tv \in C_1 \cap C_2$ for all $t > 0$. Hence $v \in T(C_1 \cap C_2; \bar{x})$, and since the other inclusion is easy, the proof is complete. □

Then, our problem becomes:

$$\mu_0 \doteq \inf\{f(x) :\ g(x) = 1,\ g_1(x) = 0,\ g_2(x) = 0\},\ g_i(x) \doteq x^\top B_i x, \tag{3.24}$$

where f and g satisfy Assumption (A).
The dual problem is defined by:

$$\nu \doteq \sup_{\lambda\in\mathbb{R}} \inf_{x\in C}\ L(\lambda, x). \tag{3.25}$$

Recall that $L(\lambda, x) \doteq f(x) + \lambda(g(x) - 1)$. Moreover, we also define the standard dual problem to (3.24):

$$\nu_S \doteq \sup_{(\lambda,\lambda_1,\lambda_2)\in\mathbb{R}^3} \inf_{x\in\mathbb{R}^n}\ L_S(\lambda, \lambda_1, \lambda_2, x), \tag{3.26}$$

where $L_S(\lambda, \lambda_1, \lambda_2, x) \doteq f(x) + \lambda(g(x) - 1) + \lambda_1 g_1(x) + \lambda_2 g_2(x)$ is termed the standard Lagrangian. We say that problem (3.24) has standard strong duality (SSD) if $\mu_0 = \nu_S$ and problem (3.26) admits solution. It is easy to check that:

$$\nu_S \le \nu \le \mu_0.$$

On the other hand, given a feasible point $\bar{x}$ ($\bar{x} \neq 0$), it is said that $\bar{x}$ is a standard KKT point to problem (3.24), if for some $(\lambda, \lambda_1, \lambda_2) \in \mathbb{R}^3$, one has the first order optimality condition:

$$\nabla f(\bar{x}) + \lambda\nabla g(\bar{x}) + \lambda_1\nabla g_1(\bar{x}) + \lambda_2\nabla g_2(\bar{x}) = 0.$$

The next result establishes a necessary and suficient optimality condition under strong duality.

Theorem 3.8. *Let f, g satisfy Assumption* (A) *with both functions being twice differentiable at the feasible point $\bar{x}$ for* (3.24)*. Let g_1, g_2 be as above. The following assertions hold.*

(*a*) *Assume that $\{B_1\bar{x}, B_2\bar{x}\}$ is linearly independent. Then $T(C;\bar{x}) = (B_1\bar{x})^\perp \cap (B_2\bar{x})^\perp$ and $(a1) \Longrightarrow (a2)$, where*

($a1$) *Strong duality holds and $\bar{x}$ is a solution for problem* (3.24)*;*

($a2$) *$\exists\ \lambda_i \in \mathbb{R}$, $i = 1, 2$, such that $\nabla f(\bar{x}) - \mu_0\dfrac{p}{q}\nabla g(\bar{x}) + \lambda_1 B_1\bar{x} + \lambda_2 B_2\bar{x} = 0$ and*

$$\nabla^2 f(\bar{x}) - \mu_0\frac{p}{q}\nabla^2 g(\bar{x}) + \lambda_1 B_1 + \lambda_2 B_2 \text{ is copositive on } T(C;\bar{x}).$$

(*b*) *Assume that $B_i\bar{x} = 0$, $i = 1, 2$. Then $T(C;\bar{x}) = C - \bar{x} = C$, and $(a1) \Longrightarrow (b1)$, where:*

($b1$) *$\nabla f(\bar{x}) - \mu_0\dfrac{p}{q}\nabla g(\bar{x}) \in [T(C;\bar{x})]^* = [T(C;\bar{x})]^\perp$ and*

$$\nabla^2 f(\bar{x}) - \mu_0\frac{p}{q}\nabla^2 g(\bar{x}) \text{ is copositive on } T(C;\bar{x}).$$

If, in addition, B_1 is positive (or negative) semidefinite, then $T(C;\bar{x}) = (\ker\ B_1)\cap(C_2 - \bar{x})$, and ($b1$) is substituted equivalently by ($b1'$):

$(b1')$ $\exists\ y \in \mathbb{R}^n$ *such that* $\nabla f(\bar{x}) - \mu_0 \dfrac{p}{q} \nabla g(\bar{x}) + B_1 y \in (C_2)^\perp$ *and*

$$\nabla^2 f(\bar{x}) - \mu_0 \frac{p}{q} \nabla^2 g(\bar{x}) \text{ is copositive on } (\ker B_1) \cap (C_2 - \bar{x}).$$

(c) *Assume that* $C = \{x :\ x^\top B_1 x = 0\}$, $B_1 \bar{x} = 0$ *and* B_1 *indefinite. Then* $T(C; \bar{x}) = C - \bar{x} = C$, $\text{co } C = \mathbb{R}^n = \text{span } C$, *and* $(a1) \Longrightarrow (c1)$, *where:*

$(c1)$ $\nabla f(\bar{x}) - \mu_0 \dfrac{p}{q} \nabla g(\bar{x}) = 0$ *and*

$$\nabla^2 f(\bar{x}) - \mu_0 \frac{p}{q} \nabla^2 g(\bar{x}) \text{ is copositive on } C - \bar{x} = C.$$

Proof. (a): $(a1) \Rightarrow (a2)$. By Proposition 3.7, and as noticed above:

$$\nabla f(\bar{x}) - \mu_0 \frac{p}{q} \nabla g(\bar{x}) \in [T(C, \bar{x})]^* = [T(C_1; \bar{x}) \cap T(C_2; \bar{x})]^* = \mathbb{R} B_1 \bar{x} + \mathbb{R} B_2 \bar{x}.$$

Hence, there exist $\lambda_i \in \mathbb{R}$, $i = 1, 2$, satisfying $\nabla_x L_S(-\mu_0 \dfrac{p}{q}, \lambda_1, \lambda_2, \bar{x}) = 0$. By Theorem 3.2, $\bar{x} \in \underset{C}{\text{argmin}}\ L(-\dfrac{p}{q}\mu_0, \cdot)$, and so $\bar{x} \in \underset{C}{\text{argmin}}\ L_S(-\dfrac{p}{q}\mu_0, \lambda_1, \lambda_2, \cdot)$.
Let $v \in T(C; \bar{x})$. Then, there exist $t_k > 0$, $x_k \in C$, $x_k \to \bar{x}$ such that $t_k(x_k - \bar{x}) \to v$. For all k sufficiently large, we obtain:

$$\begin{aligned}
0 \ &\leq\ L_S(-\mu_0 \frac{p}{q}, \lambda_1, \lambda_2, x_k) - L_S(-\mu_0 \frac{p}{q}, \lambda_1, \lambda_2, \bar{x}) \\
&=\ \nabla_x L_S(-\mu_0 \frac{p}{q}, \lambda_1, \lambda_2, \bar{x})^\top (x_k - \bar{x}) + \frac{1}{2}(x_k - \bar{x})^\top \nabla_x^2 L_S(-\mu_0 \frac{p}{q}, \lambda_1, \lambda_2, \bar{x})(x_k - \bar{x}) \\
&\quad\ + \|x_k - \bar{x}\|^2 \alpha(\bar{x}; x_k - \bar{x}) \\
&=\ \frac{1}{2}(x_k - \bar{x})^\top \nabla_x^2 L_S(-\mu_0 \frac{p}{q}, \lambda_1, \lambda_2, \bar{x})(x_k - \bar{x}) + \|x_k - \bar{x}\|^2 \alpha(\bar{x}; x_k - \bar{x}),
\end{aligned}$$

where $\alpha(\bar{x}; x - \bar{x}) \to 0$ as $x \to \bar{x}$. This implies the desired claim.
(b): $(a1) \Rightarrow (b1)$. As above, we get $\nabla f(\bar{x}) - \mu_0 \dfrac{p}{q} \nabla g(\bar{x}) \in [T(C, \bar{x})]^*$. From a previous discussion, $T(C; \bar{x}) = (C_1 - \bar{x}) \cap (C_2 - \bar{x}) = C - \bar{x} = C$, and so $[T(C; \bar{x})]^* = (C - \bar{x})^\perp = C^\perp$. Thus:

$$\nabla f(\bar{x}) - \mu_0 \frac{p}{q} \nabla g(\bar{x}) \in (C - \bar{x})^\perp.$$

Let $v \in T(C; \bar{x})$. Then, there exist $t_k > 0$, $x_k \in C$, $x_k \to \bar{x}$ such that $t_k(x_k - \bar{x}) \to v$. As in the

previous proof, it follows that for all k sufficiently large, and some $\xi \in (C-\bar{x})^{\perp}$,

$$\begin{aligned}
0 &\leq L(-\mu_0\frac{p}{q}, x_k) - L(-\mu_0\frac{p}{q}, \bar{x}) \\
&= \nabla_x L(-\mu_0\frac{p}{q}, \bar{x})^{\top}(x_k-\bar{x}) + \frac{1}{2}(x_k-\bar{x})^{\top}\nabla_x^2 L(-\mu_0\frac{p}{q}, \bar{x})(x_k-\bar{x}) \\
&\quad + \|x_k-\bar{x}\|^2\alpha(\bar{x}; x_k-\bar{x}) \\
&= \xi^{\top}(x_k-\bar{x}) + \frac{1}{2}(x_k-\bar{x})^{\top}\nabla_x^2 L(-\mu_0\frac{p}{q}, \bar{x})(x_k-\bar{x}) \\
&\quad + \|x_k-\bar{x}\|^2\alpha(\bar{x}; x_k-\bar{x}) \\
&= \frac{1}{2}(x_k-\bar{x})^{\top}\nabla_x^2 L(-\mu_0\frac{p}{q}, \bar{x})(x_k-\bar{x}) + \|x_k-\bar{x}\|^2\alpha(\bar{x}; x_k-\bar{x}). \qquad (3.27)
\end{aligned}$$

Notice that $\xi^{\top}(x_k-\bar{x}) = 0$ for all k. Thus we reach the conclusion.
The last part is a consequence of Proposition 3.6 since $C_1 = \ker B_1$.
(c): $(a1) \Rightarrow (c1)$. By the above discussion, $T(C;\bar{x}) = C$, and so $[T(C;\bar{x})]^* = (\text{co } C)^* = \{0\}$, which implies that $\nabla f(\bar{x}) - \mu_0\frac{p}{q}\nabla g(\bar{x}) = 0$. By using the same relation (3.27), one concludes that $\nabla^2 f(\bar{x}) - \mu_0\frac{p}{q}\nabla^2 g(\bar{x})$ is positive semidefinite on $C = C - \bar{x}$. □

3.4.4 The quadratic homogeneous case $p = q = 2$

We now consider the situation when:

$$f(x) \doteq \frac{1}{2}x^{\top}Ax,\ g(x) \doteq \frac{1}{2}x^{\top}Bx,$$

with A and B being real symmetric matrices, and C as in the preceding subsection.
Condition (iii) in Assumption (A) becomes:

$$x^{\top}B_1x = 0 = x^{\top}B_2x,\ x \neq 0 \Longrightarrow x^{\top}Bx > 0. \qquad (3.28)$$

This implication is a consequence of the stronger assumption:

$$\exists\ \lambda_i \in \mathbb{R}, i = 1,2,\ B + \lambda_1B_1 + \lambda_2B_2 \text{ is positive definite.}$$

We are ready to establish a necessary and sufficient optimality condition, which refines Theorem 3.8. First of all, we recall that since $p = q$, strong duality, in our sense, is satisfied.

Theorem 3.9. *Let f, g, g_1 and g_2 be as above satisfying (3.28); $\bar{x}$ be feasible for (3.24). The following assertions hold.*

- (a) *Assume that $\{B_1\bar{x}, B_2\bar{x}\}$ is linearly independent. Then $T(C;\bar{x}) = (B_1\bar{x})^{\perp} \cap (B_2\bar{x})^{\perp}$ and $(a1) \Longleftrightarrow (a2)$, where:*
  - $(a1)$ *$\bar{x}$ is a solution for problem (3.24);*
  - $(a2)$ *$\exists\ \lambda_i \in \mathbb{R}$, $i = 1,2$, such that $(A - \mu_0B + \lambda_1B_1 + \lambda_2B_2)\bar{x} = 0$ and*

$$A - \mu_0B + \lambda_1B_1 + \lambda_2B_2 \text{ is copositive on } \overline{\text{cone}}(C-\bar{x}).$$

(*b*) *Assume that* $B_i\bar{x} = 0$, $i = 1, 2$. *Then* $T(C;\bar{x}) = C - \bar{x} = C$, *and* $(a1) \Longleftrightarrow (b1)$, *where:*

$(b1)$ $(A - \mu_0 B)\bar{x} \in [T(C;\bar{x})]^* = [T(C;\bar{x})]^\perp$ *and*

$$A - \mu_0 B \text{ is copositive on } T(C;\bar{x}).$$

If, in addition, B_1 *is positive (or negative) semidefinite, then* $T(C;\bar{x}) = (\ker\, B_1) \cap (C_2 - \bar{x})$, *and* $(b1)$ *is substituted by* $(b1')$*:*

$(b1')$ $\exists\, y \in \mathbb{R}^n$ *such that* $(A - \mu_0 B)\bar{x} + B_1 y \in (C_2)^\perp$ *and*

$$A - \mu_0 B \text{ is copositive on } (\ker\, B_1) \cap (C_2 - \bar{x}).$$

Proof. (a): The first part follows from Theorem 3.8.
$(a1) \Rightarrow (a2)$. We already know that $\nabla_x L_S(-\mu_0, \lambda_1, \lambda_2, \bar{x}) = 0$ for some $\lambda_i \in \mathbb{R}$, $i = 1, 2$ and by Theorem 3.2, $\bar{x} \in \underset{C}{\operatorname{argmin}}\, L(-\mu_0, \cdot)$, and so $\bar{x} \in \underset{C}{\operatorname{argmin}}\, L_S(-\mu_0, \lambda_1, \lambda_2, \cdot)$.
For all $x \in C$, we obtain:

$$\begin{aligned} 0 &\leq L_S(-\mu_0, \lambda_1, \lambda_2, x) - L_S(-\mu_0, \lambda_1, \lambda_2, \bar{x}) \\ &= \nabla_x L_S(-\mu_0, \lambda_1, \lambda_2, \bar{x})^\top (x - \bar{x}) + \frac{1}{2}(x - \bar{x})^\top \nabla_x^2 L_S(-\mu_0, \lambda_1, \lambda_2, \bar{x})(x - \bar{x}) \\ &= \frac{1}{2}(x - \bar{x})^\top \nabla_x^2 L_S(-\mu_0, \lambda_1, \lambda_2, \bar{x})(x - \bar{x}). \end{aligned}$$

From this relation also follows the implication $(a2) \Rightarrow (a1)$.
(b): $(a1) \Leftrightarrow (b1)$. As above, we substitute x_k by any $x \in C$ in the proof of Theorem 3.8 and put $\alpha \equiv 0$.
The last part involving $(b1')$ follows easily once we notice that $C_1 = \ker\, B_1$. □

The next theorem, whose proof follows from Theorem 3.8 and Theorem 3.9, characterizes when a feasible point is optimal provided the cone C is determined by a single equality constraint: $C = \{x \in \mathbb{R}^n :\ x^\top B_1 x = 0\}$.
In this case, Condition (iii) in Assumption (A) becomes:

$$x^\top B_1 x = 0,\ x \neq 0 \Longrightarrow x^\top B x > 0.$$

It is known ([119]) that the previous implication is equivalent to:

$$\exists\, \lambda \in \mathbb{R},\ B + \lambda B_1 \text{ is positive definite.} \tag{3.29}$$

Theorem 3.10. *Let* f, g *and* g_1 *be as above satisfying* (3.29)*;* $\bar{x}$ *be feasible for* (3.24) *with* $g_2 \equiv 0$. *The following assertions hold.*

(*a*) *Assume that* $B_1\bar{x} \neq 0$. *Then* $T(C;\bar{x}) = (B_1\bar{x})^\perp$ *and* $(a1) \Longleftrightarrow (a2)$, *where*

$(a1)$ $\bar{x}$ *is a solution for problem* (3.24)*;*

$(a2)$ $\exists\, \lambda_0 \in \mathbb{R}$ *such that* $(A - \mu_0 B + \lambda_1 B_1)\bar{x} = 0$ *and*

$$A - \mu_0 B + \lambda_1 B_1 \text{ is copositive on } \overline{\text{cone}}(C - \bar{x}).$$

(b) *Assume that* $B_1\bar{x} = 0$, *and* B_1 *positive* (*or negative*) *semidefinite. Then* $(a1) \Longleftrightarrow (b1)$, *where*

$(b1)$ $\exists\ y \in \mathbb{R}^n$: $(A - \mu_0 B)\bar{x} + B_1 y = 0$ *and* $A - \mu_0 B$ *is copositive on* $\ker B_1$.

(c) *Assume that* $B_1\bar{x} = 0$, *and* B_1 *indefinite. Then* $(a1) \Longleftrightarrow (c1)$, *where*

$(c1)$ $(A - \mu_0 B)\bar{x} = 0$ *and* $A - \mu_0 B$ *is copositive on* $C - \bar{x} = C$.

3.5 Comparison between our approach and that which uses SDP relaxation

We now compare our scheme developed in previous sections and that given by the SDP relaxation as presented, for instance, in [6] and further expanded in [107]. First of all, we point out that when the SDP relaxation is successfully applied, it leads to the fulfillment of standard strong duality; whereas our approach provides a sharper lower estimate for the optimal value of the original problem in case standard strong duality fails. The comparison will be carried out for the two examples discussed in [107].

Example 3.3. (*Example 2 in [107]) Standard strong duality property fails but strong duality in our sense holds.*
Let us consider the problem:

$$\mu_0 \doteq \min\Big\{2.5x_1^2 + 2x_1x_2 + x_2^2 + 3x_3^2 :\ g(x) = 1, x \in C\Big\},\ x = (x_1, x_2, x_3), \tag{3.30}$$

where $g(x) \doteq x_1^2 + x_2^2 + x_3^2$ *and* $C \doteq \{x \in \mathbb{R}^3 :\ -2x_1x_2 \le 0,\ -x_1^2 + x_2^2 \le 0\}$. *Obviously, for these data, Assumption* (A) *is satisfied, and so by Theorem 3.1 and Corollary 3.1, we have strong duality in our sense, that is:*

$$\begin{aligned}\mu_0 = \nu &\doteq \inf_{x\in C}\{f(x) - \mu_0(g(x) - 1)\},\\ &= \max\Big\{\lambda :\ f - \lambda g \text{ is copositive on } C\Big\}.\end{aligned} \tag{3.31}$$

We already know that $\mu_0 = \nu$. *In what follows the expression* $f - \lambda g$ *will be analyzed by employing simple manipulations. More precisely, we will prove that:*

$$f(x) - \lambda g(x) \ge 0,\ \ \forall\ x \in C \Longleftrightarrow \lambda \le 2.5.$$

We split C *into three sets, and the equivalence will be checked in each of these sets.*
Case 1: $x_2 = 0$. *We obtain:* $f(x) - \lambda g(x) = (2.5 - \lambda)x_1^2 + (3 - \lambda)x_3^2$.
Case 2: $x_1 > 0, x_2 > 0$ (*so* $x_2 \le x_1$). *Clearly* $2x_2^2 \le 2x_1x_2$, *which implies:*

$$f(x) - \lambda g(x) = (2.5 - \lambda)x_1^2 + (1 - \lambda)x_2^2 + (3 - \lambda)x_3^2 + 2x_1x_2 \ge (2.5 - \lambda)x_1^2 + (3 - \lambda)(x_2^2 + x_3^2).$$

Case 3: $x_1 < 0, x_2 < 0$ (*so* $x_2 \ge x_1$). *We obtain* $2x_1x_2 \ge 2x_2^2$, *and so:*

$$f(x) - \lambda g(x) = (2.5 - \lambda)x_1^2 + (1 - \lambda)x_2^2 + (3 - \lambda)x_3^2 + 2x_1x_2 \ge (2.5 - \lambda)x_1^2 + (3 - \lambda)(x_2^2 + x_3^2).$$

By combining all the previous cases, we conclude that:

$$f(x) - \lambda g(x) \geq 0,\ \forall\ x \in C \Longleftrightarrow \lambda \leq 2.5.$$

In other words, $\mu_0 = \nu = 2.5$ *because of* (3.31)*. By looking at the objective function, one easily sees that* $\bar{x} = (\pm 1, 0, 0)$ *are two (optimal) solutions.*
What follows answers the question whether standard strong duality exists or not. To that purpose we will apply (a) *of Theorem 3.9.*
We need to re-write C *as defined by equality constraints.*
Set now $x = (x_1, x_2, x_3, x_4, x_5)$*,* $g_1(x) \doteq -2x_1x_2 + x_4^2$*,* $g_2(x) \doteq -x_1^2 + x_2^2 + x_5^2$*, so that (we keep the same notation* C *for the conical constraint):*

$$C \doteq \{x \in \mathbb{R}^5 :\ -2x_1x_2 + x_4^2 = 0,\ -x_1^2 + x_2^2 + x_5^2 = 0\}.$$

We use the same terminology as in Subsection 3.4.4. Here, the solutions become $\bar{x} = (\pm 1, 0, 0, 0, 0)$*. By identifying the matrices* A*,* B*,* B_1, B_2*, one gets* $\{B_1\bar{x}, B_2\bar{x}\}$ *is linearly independent (so* (a) *of Theorem 3.9 is applicable) and*

$$(A - 2.5B + \lambda_1 B_1 + \lambda_2 B_2)\bar{x} = 0 \Longleftrightarrow \lambda_1 = 1, \lambda_2 = 0.$$

Simple computations show that there is no $\lambda_i \in \mathbb{R}$*,* $i = 1, 2$*, such that* $A - 2.5B + \lambda_1 B_1 + \lambda_2 B_2$ *be positive semidefinite, and so standard strong duality fails. But our Theorem 3.9 provides the extra information that* $A - 2.5B + B_1$ *is copositive on* $\overline{\text{cone}}(C - \bar{x})$*.*

Under SDP relaxation
This approach requires the notation, given any two matrices of the same size U *and* V*,*

$$U \bullet V \doteq \text{trace}(UV^\top) = \text{tr}(UV^\top),$$

and uses the SDP relaxation to Problem (3.30) *defined as:*

$$\begin{array}{rll} \bar{\mu} = \min & Q_0 \bullet X & \\ st & Q_i \bullet X \leq 0,\ i = 1, 2 & \\ & I \bullet X = 1,\ X \succeq 0, & \end{array} \tag{3.32}$$

where:

$$Q_0 = \begin{bmatrix} 2.5 & 1 & 0 \\ 1 & 1 & 0 \\ 0 & 0 & 3 \end{bmatrix},\ Q_1 = \begin{bmatrix} 0 & -1 & 0 \\ -1 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix},\ Q_2 = \begin{bmatrix} -1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix},\ X = xx^\top.$$

The (conical) dual problem to (3.32) *is:*

$$\begin{array}{rll} \bar{\nu} = \sup & -\lambda & \\ st & Z = Q_0 + \mu_1 Q_1 + \mu_2 Q_2 + \lambda I \succeq 0, & \\ & \mu_1 \geq 0,\ \mu_2 \geq 0. & \end{array} \tag{3.33}$$

Obviously, $\bar{\nu} \leq \bar{\mu} \leq \mu_0$. One can easily see that Problem (3.33) *satisfies the Slater condition, i.e, there exist $\mu_i \geq 0$, $i = 1, 2$, $\lambda \in \mathbb{R}$ such that Z is positive definite* (*it is possible by choosing λ sufficiently large*). *On the other hand, it is shown in [107] that $\tilde{x} = \left(2/\sqrt{5}, 1/\sqrt{5}, 0\right)$ is a Slater point of Problem* (3.30) ($g(\tilde{x}) = 1$, $\tilde{x}^\top Q_i \tilde{x} < 0$, $i = 1, 2$), *which implies that its SDP relaxation* (3.32) *also admits a Slater point. As a consequence, $\bar{\mu} = \bar{\nu}$ and both relaxed problems,* (3.32) *and* (3.33), *admit solutions. In addition, it is known that, primal-dual feasible pair $\bar{X}$ and $(\bar{Z}, \bar{\lambda}, \bar{\mu}_1, \bar{\mu}_2)$ are optimal if, and only if they satisfy the complementary conditions:*

$$\bar{Z}\bar{X} = 0,\ \bar{\mu}_1 Q_1 \bullet \bar{X} = 0,\ \bar{\mu}_2 Q_2 \bullet \bar{X} = 0.$$

In [107], the authors solve problems (3.32) *and* (3.33), *and find the unique solutions, respectively,*

$$\bar{X} = \begin{bmatrix} 0.5 & 0 & 0 \\ 0 & 0.5 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \quad \textit{and}$$

$$\bar{Z} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1.25 \end{bmatrix},\ \bar{\lambda} = -1.75,\ \bar{\mu}_1 = 0.75,\ \bar{\mu}_2 = 1.$$

In order to verify whether there exists or not tightness, which means $\mu_0 = \bar{\mu}$, one applies Theorem 1 in [107]. This result ensures that (3.32) *is tight if, and only if "Property* J*"* (*as defined in the same work*) *fails.*
We recall that an optimal complementary pair $\bar{X}$ and $(\bar{Z}, \bar{\lambda}, \bar{\mu}_1, \bar{\mu}_2)$ is said to have Property J, *if the following conditions are simultaneously satisfied*:

1. $\bar{\mu}_1 \bar{\mu}_2 > 0$;
2. rank $\bar{Z} = n - 2$;
3. rank $\bar{X} = 2$, *and there is a rank-one decomposition of $\bar{X}$ as $\bar{X} = x^1 {x^1}^\top + x^2 {x^2}^\top$, such that:*
   3.1. $Q_1 \bullet x^1 {x^1}^\top = Q_1 \bullet x^2 {x^2}^\top = 0$, $[Q_2 \bullet x^1 {x^1}^\top][Q_2 \bullet x^2 {x^2}^\top] < 0]$;
   3.2. ${x^1}^\top Q_1 x^2 \neq 0$. *That is, x^1 and x^2 are not Q_1-orthogonal.*

In the present situation, the three conditions required by Property J *are satisfied, as it is checked in [107]. Thus, $1.75 = \bar{\nu} = \bar{\mu} < \mu_0$. Observe that*

$$f(2/\sqrt{5}, 1/\sqrt{5}, 0) = 3,\ f(1/\sqrt{2}, 1/\sqrt{2}, 0) = 2.75.$$

Here, $(1/\sqrt{2}, 1/\sqrt{2}, 0)$ is obtained after using a rank-one decomposition to the above matrix $\bar{X}$. Finally, the original problem (3.30) *is solved once we substitute $x_3^2 = 1 - x_1^2 - x_2^2$ on it. Thus, one finds that*

$$\underset{K}{\operatorname{argmin}}\ f = \{(\pm 1, 0, 0)\}.$$

Example 3.4. (*Example 3 in [107]*) *Standard strong duality holds.*
Let us consider the problem:

$$\mu_0 \doteq \min\left\{2x_1^2 - 2x_2x_3 + 20x_3^2 :\ g(x) = 1, x \in C\right\},\ x = (x_1, x_2, x_3), \tag{3.34}$$

where $g(x) \doteq x_1^2 + x_2^2 + x_3^2$ *and* $C \doteq \{x \in \mathbb{R}^3 : \ 2x_2x_3 + x_3^2 \leq 0, \ x_1^2 - x_2^2 \leq 0\}$. *As previously, Assumption* (A) *is satisfied, and so by Theorem 3.1 and Corollary 3.1, we have strong duality:*

$$\begin{aligned} \mu_0 = \nu &\doteq \inf_{x \in C} \{f(x) - \mu_0 \lambda(g(x) - 1)\}, \\ &= \max\{\lambda : \ f - \lambda g \text{ is copositive on } C\Big\}. \end{aligned} \tag{3.35}$$

Thus $\bar{\lambda} = \mu_0 = \nu$. *We analyze when the inequality* $f - \lambda g \geq 0$ *holds.*
We re-write C *as* $C \doteq \{x \in \mathbb{R}^3 : \max\{(x_2 + x_3)^2, x_1^2\} \leq x_2^2\}$. *Since* $\max\{a, b\} = (a + b + |a - b|)/2$, *we obtain*

$$\begin{aligned} f(x) - \lambda g(x) &= -(x_2 + x_3)^2 + 3x_2^2 + 21x_3^2 - \lambda(x_1^2 + x_2^2 + x_3^2) \\ &\geq -2x_2^2 + x_1^2 + |(x_2 + x_3)^2 - x_1^2| + 3x_2^2 + 21x_3^2 - \lambda(x_1^2 + x_2^2 + x_3^2) \\ &\geq (1 - \lambda)x_1^2 + (1 - \lambda)x_2^2 + (21 - \lambda)x_3^2 + |(x_2 + x_3)^2 - x_1^2|. \end{aligned}$$

This implies that if $1 - \lambda \geq 0$ *then* $f(x) - \lambda g(x) \geq 0$ *for all* $x \in C$. *Hence, from* (3.35) *it follows that* $\mu_0 = \nu \geq 1$. *By looking at the objective function, we immediately realize that* $f\left(\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}, 0\right) = 1$, *and therefore* $\mu_0 = \nu = 1$. *Consequently*

$$\operatorname*{argmin}_K f = \left\{ \left(\pm \frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}, 0 \right), \left(\pm \frac{1}{\sqrt{2}}, -\frac{1}{\sqrt{2}}, 0 \right) \right\}.$$

It remains only to see whether there exists or not standard strong duality. We will apply again (a) *of Theorem 3.9.*
By re-writing the constraints by equalities as in the previous example by adding the variables x_4 *and* x_5. *Set now* $x \doteq (x_1, x_2, x_3, x_4, x_5)$, $g_1(x) \doteq 2x_2x_3 + x_3^2 + x_4^2$, $g_2(x) \doteq x_1^2 - x_2^2 + x_5^2$, *so that (use the same notation for* C*)* $C \doteq \{x \in \mathbb{R}^5 : \ g_1(x) = 0, \ g_2(x) = 0\}$.
By identifying the matrices A, B, B_1, B_2, *we obtain for all* $\bar{x} \in \operatorname*{argmin}_K f$*:*

$$(A - B + \lambda_1 B_1 + \lambda_2 B_2)\bar{x} = 0 \Longleftrightarrow \lambda_1 = \lambda_2 = 1.$$

Then $A - B + B_1 + B_2$ *is positive semidefinite, which yields standard strong duality.*

Under SDP relaxation
This approach uses the SDP relaxation to Problem (3.34). *As in the previous case, the authors in [107] find that* $\bar{x} = \left(0, -\frac{2}{\sqrt{5}}, \frac{1}{\sqrt{5}}\right)$ *is a Slater point, and by direct computations, obtain an optimal complementary pair*

$$\bar{X} = \begin{bmatrix} 0.5 & 0 & 0 \\ 0 & 0.5 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \quad \textit{and}$$

$$\bar{Z} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 20 \end{bmatrix}, \ \bar{\lambda} = 1 \ \bar{\mu}_1 = \bar{\mu}_2 = 1.$$

Thus $\bar{\mu} = \bar{\nu} = 1$. *From this, the authors in [107] check that condition 3.2 in Property* J *fails. Hence, there is tightness, and so standard strong duality is satisfied. By using the rank-one decomposition method, from* $\bar{X}$, *one gets an optimal solution given by* $\bar{x} = \left(\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}, 0\right)$.

3.6 A class of quadratic fractional optimization problems with two quadratic constraints

Motivated by applications, we now consider the problem of minimizing an indefinite quadratic fractional function under two quadratic constraints. This problem was also studied in [52] via the approach of SDP relaxation, so the authors provide conditions under which, as a consequence, standard strong duality is satisfied. On the contrary, our study goes beyond this situation. The problem is the following:

$$\mu_{qf} \doteq \min \Big\{ \frac{w^\top A w + a^\top x + \alpha}{w^\top B w + b^\top w + \beta} : \ w^\top B_1 w + b_1^\top w + \beta_1 \leq 0, \ w^\top w - \delta \leq 0 \Big\}, \tag{3.36}$$

where $\delta > 0$; a, b, b_1 are in $\mathbb{R}^n$, $n \geq 2$; $\alpha, \beta, \beta_1 \in \mathbb{R}$; and A, B, B_1 are (real) symmetric matrices of order n. The feasible set of problem (3.36) is:

$$K_0 \doteq \{w \in \mathbb{R}^n : \ w^\top B_1 w + b_1^\top w + \beta_1 \leq 0, \ w^\top w - \delta \leq 0\}.$$

A natural assumption to be imposed is:

$$w^\top B w + b^\top w + \beta > 0, \quad \forall \ w \in K_0. \tag{3.37}$$

By using the generalized Charnes-Cooper transformation:

$$z = \frac{1}{\sqrt{w^\top B w + b^\top w + \beta}}, \ y = \frac{w}{\sqrt{w^\top B w + b^\top w + \beta}}, \tag{3.38}$$

problem (3.36) can be written "equivalently" (to be precised later) to the following homogeneous quadratic minimization problem:

$$\begin{array}{rll} \mu_0 \doteq \min & y^\top A y + a^\top y z + \alpha z^2 & \\ \text{st} & y^\top B y + b^\top y z + \beta z^2 = 1, & \\ & y^\top B_1 y + b_1^\top y z + \beta_1 z^2 \leq 0, & \\ & y^\top y - \delta z^2 \leq 0, & \end{array} \tag{3.39}$$

which fits our model with $p = q = 2$, impliying the validity of strong duality for (3.39), since $\delta > 0$ and under (3.37), Assumption (A) is fulfilled for (3.39).
Problems (3.39) and (3.36) are equivalent in the sense that:

- if (y, z) is feasible (optimal) for problem (3.39) then $z \neq 0$ and $w \doteq y/z$ is feasible (optimal) for (3.36);
- if w is feasible (optimal) for problem (3.36) then (y, z) given by (3.38) is feasible (optimal) for (3.39).

Hence, $\mu_{qf} = \mu_0$, and so solutions to (3.36) can be obtained by solving a nonconvex homogeneous quadratic problem.
In order to become the inequality constraints in problem (3.39) into equalities, we introduce the slack variables x_{n+2}, x_{n+3}. Thus, setting $x \doteq (y, z, x_{n+2}, x_{n+3})$, along with:

$$f(x) \doteq x^\top \bar{A} x,\ g(x) \doteq x^\top \bar{B} x,\ g_1(x) \doteq x^\top \bar{B}_1 x,\ g_2(x) \doteq x^\top \bar{B}_2 x,$$

where:

$$\bar{A} = \begin{pmatrix} A & \frac{a}{2} & 0 & 0 \\ \frac{a^\top}{2} & \alpha & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix},\ \bar{B} = \begin{pmatrix} B & \frac{b}{2} & 0 & 0 \\ \frac{b^\top}{2} & \beta & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix},$$

$$\bar{B}_1 = \begin{pmatrix} B_1 & \frac{b_1}{2} & 0 & 0 \\ \frac{b_1^\top}{2} & \beta_1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix},\ \bar{B}_2 = \begin{pmatrix} I & 0 & 0 & 0 \\ 0 & -\delta & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix},$$

problem (3.36) reduces

$$\mu_0 = \min\{f(x) :\ g(x) = 1,\ x \in C\}. \tag{3.40}$$

Here, $C \doteq \{x \in \mathbb{R}^{n+3} :\ g_1(x) = 0,\ g_2(x) = 0\}$. As above, if $\delta > 0$ and (3.37) holds, then f, g and C satisfy Assumption (A). To make more readable the expressions, let us introduce some notation:

$$C_0 \doteq \{(y,z) \in \mathbb{R}^{n+1} : g_1(y,z,0,0) \le 0,\ g_2(y,z,0,0) \le 0\};$$

$$\tilde{f}_0(w) \doteq w^\top A w + a^\top w + \alpha;\ \tilde{g}_0(w) \doteq w^\top B w + b^\top w + \beta;\ h(w) \doteq \frac{\tilde{f}_0(w)}{\tilde{g}_0(w)}.$$

$$f_0(y,z) \doteq z^2 \tilde{f}_0\Big(\frac{y}{z}\Big);\ g_0(y,z) \doteq z^2 \tilde{g}_0\Big(\frac{y}{z}\Big) = g(y,z,0,0);$$

$$\tilde{g}_1(w) \doteq w^\top B_1 w + b_1^\top w + \beta_1 = g_1(w,1,0,0);\ \tilde{g}_2(w) \doteq w^\top w - \delta = g_2(w,1,0,0).$$

The following remark will be taken into account without mentioning it explicitly.

Remark 3.8. *For the data as above:*

- *if $x = (y, z, x_{n+2}, x_{n+3})$ is feasible for problem (3.40) then $z \neq 0$ and $w \doteq y/z \in K_0$;*
- *if $w \in K_0$ then:*

$$\frac{1}{\sqrt{\tilde{g}_0(w)}}\Big(w, 1, \pm\sqrt{-\tilde{g}_1(w)}, \pm\sqrt{-\tilde{g}_2(w)}\Big) \text{ is feasible for (3.40).}$$

- $\operatorname{cone}(C - tx) = \operatorname{cone}(C - x)$ *for all $x \in C$ and all $t > 0$.*
- *the following hold:*

$$\begin{aligned} (y,z,x_{n+2},x_{n+3}) \in C &\iff (-y,-z,x_{n+2},x_{n+3}) \in C \\ &\iff (y,z,-x_{n+2},x_{n+3}) \in C \\ &\iff (-y,-z,-x_{n+2},-x_{n+3}) \in C. \end{aligned} \tag{3.41}$$

By virtue of Theorem 3.9, we need to compute the gradient of g_1 and g_2. It is easily seen that:

$$\nabla g_1(x) = 2\begin{pmatrix} B_1 y + \frac{z}{2} b_1 \\ \frac{1}{2} b_1^\top y + \beta_1 z \\ x_{n+2} \\ 0 \end{pmatrix}, \quad \nabla g_2(x) = 2\begin{pmatrix} y \\ -\delta z \\ 0 \\ x_{n+3} \end{pmatrix} \text{ with } x = (y, z, x_{n+2}, x_{n+3}). \tag{3.42}$$

The next lemma provides very mild conditions under which linear independence of the gradients is obtained.

Lemma 3.3. *Let us consider the above data and assume that $x = (y, z, x_{n+2}, x_{n+3})$ is feasible for problem* (3.40)*. The following assertions hold:*

(*a*) $\nabla g_2(x) \neq 0$.

(*b*) *If either $x_{n+2} \neq 0$ or $x_{n+3} \neq 0$, then $\{\nabla g_1(x), \nabla g_2(x)\}$ is linearly independent.*

(*c*) *If $x_{n+2} = 0 = x_{n+3}$, then*

(*c*1) *$\{\nabla g_1(x), \nabla g_2(x)\}$ is linearly dependent $\Longleftrightarrow \nabla g_1(x) = \gamma \nabla g_2(x)$ for some $\gamma \in \mathbb{R}$;*

(*c*2) *$\{\nabla g_1(x), \nabla g_2(x)\}$ is linearly independent $\Longleftrightarrow \left\{B_1 y + \frac{z}{2} b_1, y\right\}$ is linearly independent.*

Proof. (*a*) This follows from the feasibility of x; (*b*) is a consequence of (3.42).
(*c*1) is straighforward; (*c*2): By (3.42) again, we get $\{\nabla g_1(x), \nabla g_2(x)\}$ is linearly independent if, and only if

$$\left\{ \begin{pmatrix} B_1 y + \frac{z}{2} b_1 \\ \frac{1}{2} b_1^\top y + \beta_1 z \end{pmatrix}, \begin{pmatrix} y \\ -\delta z \end{pmatrix} \right\} \text{ is linearly independent.}$$

This is equivalent to the linear independence of $\left\{B_1 y + \frac{z}{2} b_1, y\right\}$ by feasibility again. □

Concerning (*c*) of the preceding lemma, it is not hard to check the equivalence between (3.43) and (3.44) relating the variable x and the original one w:

$$[g(y, z, 0, 0) = 1,\ g_1(y, z, 0, 0) = 0, g_2(y, z, 0, 0) = 0] \Longrightarrow \left\{B_1 y + \frac{z}{2} b_1, y\right\} \text{ is LI.} \tag{3.43}$$

$$[g_1(w, 1, 0, 0) = 0,\ g_2(w, 1, 0, 0) = 0] \Longrightarrow \left\{B_1 w + \frac{1}{2} b_1, w\right\} \text{ is LI.} \tag{3.44}$$

We are ready to establish a characterization of optimality under linear independence of the gradients.

Theorem 3.11. *Let us consider the above data and $\bar{w}$ be feasible for problem* (3.36)*. Assume that any of the conditions* (*a*)*,* (*b*)*, or* (*c*) *is satisfied:*

(*a*) $\bar{w}^\top \bar{w} < \delta$;

(*b*) $\bar{w}^\top B_1 \bar{w} + b_1^\top \bar{w} + \beta_1 < 0$;

(*c*) $\bar{w}^\top \bar{w} = \delta$, $\bar{w}^\top B_1 \bar{w} + b_1^\top \bar{w} + \beta_1 = 0$ *and* $\left\{ B_1 \bar{w} + \frac{1}{2} b_1, \bar{w} \right\}$ *is linearly independent.*

Then, $\bar{w}$ *is optimal to* (3.36) *if, and only if there exist* $\lambda_1, \lambda_2 \in \mathbb{R}$ *such that:*

$$\begin{aligned} A\bar{w} + \frac{1}{2}a - \mu_0 \left(B\bar{w} + \frac{1}{2}b \right) + \lambda_1 \left(B_1 \bar{w} + \frac{1}{2} b_1 \right) + \lambda_2 \bar{w} = 0 \\ \frac{1}{2} a^\top \bar{w} + \alpha - \mu_0 \left(\frac{1}{2} b^\top \bar{w} + \beta \right) + \lambda_1 \left(\frac{1}{2} b_1^\top \bar{w} + \beta_1 \right) - \lambda_2 \delta = 0 \\ \lambda_1 (\bar{w}^\top B_1 \bar{w} + b_1^\top \bar{w} + \beta_1) = 0 \\ \lambda_2 (\bar{w}^\top \bar{w} - \delta) = 0 \end{aligned}$$

and $\bar{A} - \mu_0 \bar{B} + \lambda_1 \bar{B}_1 + \lambda_2 \bar{B}_2$ *is copositive on* $\overline{\text{cone}}(C - \bar{x})$*, where* $\bar{x}$ *belongs to*

$$\frac{1}{\sqrt{\tilde{g}_0(\bar{w})}} \Big\{ (\bar{w}, 1, \tilde{x}_{n+2}, \tilde{x}_{n+3}), (-\bar{w}, -1, \tilde{x}_{n+2}, \tilde{x}_{n+3}), (\bar{w}, 1, -\tilde{x}_{n+2}, \tilde{x}_{n+3}),$$

$$(-\bar{w}, -1, -\tilde{x}_{n+2}, -\tilde{x}_{n+3}) \Big\},$$

with:

$$\tilde{x}_{n+2} = \sqrt{-\bar{w}^\top B_1 \bar{w} - b_1^\top \bar{w} - \beta_1} \quad \textit{and} \quad \tilde{x}_{n+3} = \sqrt{-\bar{w}^\top \bar{w} + \delta}.$$

Proof. $\Rightarrow$: Let $\bar{w}$ be optimal to (3.36). Obviously there exist $\tilde{x}_{n+2}, \tilde{x}_{n+3} \in \mathbb{R}$ satisfying $\bar{w}^\top B_1 \bar{w} + b_1^\top \bar{w} + \beta_1 + \tilde{x}_{n+2}^2 = 0, \bar{w}^\top \bar{w} - \delta + \tilde{x}_{n+3}^2 = 0$. This allows us to infer, in view of (3.41), that any $\bar{x}$ belonging to:

$$\frac{1}{\sqrt{\tilde{g}_0(\bar{w})}} \Big\{ (\bar{w}, 1, \tilde{x}_{n+2}, \tilde{x}_{n+3}), (\bar{w}, 1, -\tilde{x}_{n+2}, \tilde{x}_{n+3}), (\bar{w}, 1, \tilde{x}_{n+2}, -\tilde{x}_{n+3}), (-\bar{w}, -1, \tilde{x}_{n+2}, -\tilde{x}_{n+3}) \Big\}$$

is optimal for problem (3.40) and $\mu_{qf} = \mu_0$, as noticed above. Thus, by Lemma 3.3 and Theorem 3.9 (*a*), the conclusion is obtained. □

In case of active feasible points, the preceding result becomes more precise. It asserts that solving problem (3.36) amounts to solve a linear system with $n+1$ equations and three unknown variables; where the matrix defining such a system has linearly independent column vectors.

Theorem 3.12. *Let us consider the above data. Assume that* $\bar{w}^\top \bar{w} = \delta$, $\bar{w}^\top B_1 \bar{w} + b_1^\top \bar{w} + \beta_1 = 0$ *and* $\left\{ B_1 \bar{w} + \frac{1}{2} b_1, \bar{w} \right\}$ *is linearly independent. Then, the following hold:*

(*a*) $\bar{w}$ *is optimal to problem* (3.36) *if, and only if there exist* $\lambda_1, \lambda_2 \in \mathbb{R}$ *such that:*

$$\left(-B\bar{w} - \frac{1}{2} b \right) \mu_0 + \left(B_1 \bar{w} + \frac{1}{2} b_1 \right) \lambda_1 + \bar{w} \lambda_2 = -A\bar{w} - \frac{1}{2} a \tag{3.45}$$

$$\left(-\frac{1}{2} b^\top \bar{w} - \beta \right) \mu_0 + \left(\frac{1}{2} b_1^\top \bar{w} + \beta_1 \right) \lambda_1 - \delta \lambda_2 = -\frac{1}{2} a^\top \bar{w} - \alpha. \tag{3.46}$$

(*b*) *The matrix defining the linear system* (3.45)-(3.46) *has linearly independent columns.*

Proof. (a): The first order optimality condition follows from the preceding theorem. Conversely, by combining (3.45) and (3.46), we get $\bar{w}^\top A\bar{w}+a^\top\bar{w}+\alpha-\mu_0(\bar{w}^\top B\bar{w}+b^\top\bar{w}+\beta)=0$, which means that $\bar{w}$ is optimal.
(b): Let h_i, $i=1,2,3$ be the columns of the matrix, $H\in\mathbb{R}^{(n+1)\times 3}$, determining the linear system (3.45)-(3.46), and let d be its right-hand side vector, that is,

$$H=(h_1\ \ h_2\ \ h_3)=\begin{pmatrix}-B\bar{w}-\frac{1}{2}b & B_1\bar{w}+\frac{1}{2}b_1 & \bar{w}\\ -\frac{1}{2}b^\top\bar{w}-\beta & \frac{1}{2}b_1^\top\bar{w}+\beta_1 & -\delta\end{pmatrix},\ d=\begin{pmatrix}-A\bar{w}-\frac{1}{2}a\\ -\frac{1}{2}a^\top\bar{w}-\alpha\end{pmatrix}.$$

Suppose to the contrary that $\{h_1,h_2,h_3\}$ is linearly dependent. Then, there exist $(\gamma_1,\gamma_2,\gamma_3)\in\mathbb{R}^3\setminus\{0\}$ such that:

$$\gamma_1h_1+\gamma_2h_2+\gamma_3h_3=0. \tag{3.47}$$

By assumption, $\{h_2,h_3\}$ is linearly independent, therefore $\gamma_1\neq 0$. Without loss of generality, we may suppose that $\gamma_1=1$. Thus (3.47) reduces:

$$-B\bar{w}-\frac{1}{2}b+\gamma_2\left(B_1\bar{w}+\frac{1}{2}b_1\right)+\gamma_3\bar{w}=0 \tag{3.48}$$

$$-\frac{1}{2}b^\top\bar{w}-\beta+\gamma_2\left(\frac{1}{2}b_1^\top\bar{w}+\beta_1\right)-\gamma_3\delta=0. \tag{3.49}$$

From (3.49) it follows that:

$$\gamma_3=\frac{-b^\top\bar{w}-2\beta+\gamma_2\left(b_1^\top\bar{w}+2\beta_1\right)}{2\delta}.$$

By substituting in (3.48), we get:

$$-2B\bar{w}-b+\gamma_2\left(2B_1\bar{w}+b_1\right)+\left(\frac{-b^\top\bar{w}-2\beta+\gamma_2\left(b_1^\top\bar{w}+2\beta_1\right)}{\delta}\right)\bar{w}=0.$$

Taking the product with $\bar{w}^\top$ and using the facts that $\bar{w}^\top\bar{w}=\delta$ and $\bar{w}^\top B_1\bar{w}+b_1^\top\bar{w}+\beta_1=0$, we obtain:

$$-2\bar{w}^\top B\bar{w}-b^\top\bar{w}+\gamma_2\left(2\bar{w}^\top B_1\bar{w}+b_1^\top\bar{w}+b_1^\top\bar{w}+2\beta_1\right)-b^\top\bar{w}-2\beta=0,$$

which implies $\bar{w}^\top B\bar{w}+b^\top\bar{w}+\beta=0$. This is impossible, proving that H has linearly independent columns, and so rank $H=3$. □

Part (b) of Theorem 3.12 allows us to employ the (Moore-Penrose) pseudoinverse $H^\dagger$ of H for determining a solution $(\mu_0,\lambda_1,\lambda_2)$. Indeed, since H has full column rank, it is known that $H^\dagger=(H^\top H)^{-1}H^\top$ and $H=QR$, where $Q\in\mathbb{R}^{(n+1)\times 3}$ has orthonormal columns ($Q^\top Q=I$) and $R\in\mathbb{R}^{3\times 3}$ is upper triangular with non-null diagonal elements. Thus, combining both results, we have that $H^\dagger=R^{-1}Q^\top$, see [20, 134]. Therefore, system (3.45)-(3.46) admits a unique solution given by $H^\dagger d$ if, and only if $HH^\dagger d=d$ if, and only if $QQ^\top d=d$.
By writting:

$$R = \begin{pmatrix} R_{11} & R_{12} & R_{13} \\ 0 & R_{22} & R_{23} \\ 0 & 0 & R_{33} \end{pmatrix}, \quad R^{-1} = \begin{pmatrix} r_{11} & r_{12} & r_{13} \\ 0 & r_{22} & r_{23} \\ 0 & 0 & r_{33} \end{pmatrix},$$

and solving the following system of equations:

$$\begin{cases} r_{11}R_{12} + r_{12}R_{22} = 0 \\ r_{11}R_{13} + r_{12}R_{23} + r_{13}R_{33} = 0 \\ r_{22}R_{23} + r_{23}R_{33} = 0 \\ r_{ii} = R_{ii}^{-1} \ i = 1, 2, 3, \end{cases}$$

we get the expression for R^{-1}. Hence, the unique solution of the linear system (under consistency) is:

$$\mu_0 = \left(\frac{q_1}{R_{11}} - \frac{R_{12}}{R_{11}R_{22}} q_2 + \left[\frac{R_{12}R_{23}}{R_{11}R_{22}R_{33}} - \frac{R_{13}}{R_{11}R_{33}} \right] q_3 \right)^\top d \tag{3.50a}$$

$$\lambda_1 = \left(\frac{q_2}{R_{22}} - \frac{R_{23}}{R_{22}R_{33}} q_3 \right)^\top d \tag{3.50b}$$

$$\lambda_2 = \frac{q_3^\top d}{R_{33}}, \tag{3.50c}$$

where q_1, q_2, q_3 are the three columns of Q. Notice that all the data required in (3.50) are readily available through few steps of the classical Gram-Schmidt algorithm [20]:

G-S Algorithm

for $k = 1 : 3$

  for $i = 1 : k - 1$

    $R_{ik} = q_i^\top h_k$;

  end

  $\hat{q}_k = h_k - \sum_{i=1}^{k-1} R_{ik} q_i$;

  $R_{kk} = ||\hat{q}_k||$;

  $q_k = \hat{q}_k / R_{kk}$

end

In general, even if the linear system is not consistent, $R^{-1}Q^\top d$ minimizes the function $\xi \mapsto \|H\xi - d\|_2$ over $\xi \in \mathbb{R}^3$.

3.7 L-eigenvalues as an extension of H and Z-eigenvalues in real symmetric tensors problems

Let $m, n \in \mathbb{N}$ with $m \geq 2$, $n \geq 2$. A real m-order n-dimensional tensor $\mathcal{A}$ consists of n^m entries in $\mathbb{R}$, and it is denoted by:

$$\mathcal{A} = (\mathcal{A}_{i_1 i_2 \ldots i_m}),\ i_1, i_2, \ldots, i_m \in \{1, 2, \ldots, n\}.$$

We say a tensor $\mathcal{A}$ is symmetric (the term supersymmetric is also used by some authors) if its entries $\mathcal{A}_{i_1 i_2 \ldots i_m}$ are invariant under any permutation of the indices $(i_1 i_2 \ldots i_m)$. Furthermore, given any $x \in \mathbb{R}^n$, it is defined:

$$\mathcal{A}x^m \doteq \sum_{i_1, \cdots, i_m = 1}^{n} \mathcal{A}_{i_1 i_2 \cdots i_m} x_{i_1} \ldots x_{i_m},$$

which is an mth-degree homogeneous polynomial whenever $\mathcal{A}$ is symmetric. Throughout this section the tensor $\mathcal{A}$ will be symmetric.
Given $x \in \mathbb{R}^n$: $\|x\|_k$ stands for the l^k-norm; $\mathcal{A}x^{m-1}$ is the vector in $\mathbb{R}^n$ whose i-th component is,

$$\left(\mathcal{A}x^{m-1}\right)_i = \sum_{i_2, \cdots, i_m = 1}^{n} \mathcal{A}_{i\ i_2 \cdots i_m} x_{i_2} \ldots x_{i_m} \quad \text{for } i = 1, \ldots, n,$$

and set $x^{[k]} \doteq (x_1^k, \ldots, x_n^k)$ and $x^{[0]} \doteq (1, \ldots, 1)$.
We now consider the following constrained optimization problem:

$$\mu_k \doteq \min \left\{ \mathcal{A}x^m :\ \|x\|_k^m = 1, \quad x \in C \right\}, \tag{3.51}$$

where C is a closed cone in $\mathbb{R}^n$. Clearly, under symmetry on $\mathcal{A}$, problem (3.51) is a particular model of (3.1), where:

$$f(x) = \mathcal{A}x^m, \quad g(x) = \|x\|_k^m. \tag{3.52}$$

Both functions f and g have the same degree of homogeneity, m, and satisfy Assumption (A). Thus (Theorem 3.1):

$$\mu_k = \nu_k \doteq \sup_{\lambda \in \mathbb{R}} \inf_{x \in C} \left\{ \mathcal{A}x^m + \lambda(\|x\|_k^m - 1) \right\} = \inf_{x \in C} \left\{ \mathcal{A}x^m - \mu_k(\|x\|_k^m - 1) \right\}.$$

By Theorem 3.2, $\bar{x}$ is a solution to (3.51) if and only if:

$$\|\bar{x}\|_k^m = 1 \ \text{ and } \ \bar{x} \in \operatorname*{argmin}_C \ L(-f(\bar{x}), \cdot).$$

Hence, in order to obtain first- and second-order necessary optimality conditions, we can use Theorem 3.3 or Theorem 3.8 (when C is determined by two quadratic forms); whereas Theorem 3.4 provides sufficient optimality conditions for strict local optimality. Here, we need the following computation, given any $x \in C$, $x \neq 0$,

$$\nabla f(x) = m\mathcal{A}x^{m-1},\ \nabla^2 f(x) = m(m-1)\mathcal{A}x^{m-2}.$$

The symbol $\mathcal{A}x^{m-2}$ denotes the $(m-2)$-times product of the tensor $\mathcal{A}$ with the vector x, which is defined as the matrix of $\mathbb{R}^{n \times n}$ whose entries are

$$(\mathcal{A}x^{m-2})_{i_1 i_2} = \sum_{i_3, \ldots, i_m = 1}^{n} A_{i_1 i_2 i_3 \ldots i_m} x_{i_3} \cdots x_{i_m}.$$

Moreover,

$$\nabla g(x) = m\|x\|_k^{m-k}\varphi_k(x), \quad \nabla^2 g(x) = \Big(\nabla^2 g(x)_{ij}\Big),$$

where $\varphi_k(x) \doteq (x_1|x_1|^{k-2}, x_2|x_2|^{k-2}, \ldots, x_n|x_n|^{k-2})$ and

$$\nabla^2 g(x)_{ij} = \begin{cases} m(m-k)\|x\|_k^{m-2k}(\varphi_k(x)_i)^2 + m\|x\|_k^{m-k}(k-1)|x_i|^{k-2}, \text{ if } i = j; \\ \\ m(m-k)\|x\|_k^{m-2k}\varphi_k(x)_i\varphi_k(x)_j, \text{ if } i \neq j. \end{cases}$$

Here, $\varphi_k(x)_i$ stands for the i-th component of the vector $\varphi_k(x)$.

Observe that in case $C \subseteq \mathbb{R}^n_+$, one gets $\nabla g(x) = m\|x\|_k^{m-k}x^{[k-1]}$ and

$$\nabla^2 g(x)_{ij} = \begin{cases} m(m-k)\|x\|_k^{m-2k}(x_i^{k-1})^2 + m\|x\|_k^{m-k}(k-1)x_i^{k-2}, \text{ if } i = j; \\ \\ m(m-k)\|x\|_k^{m-2k}x_i^{k-1}x_j^{k-1}, \text{ if } i \neq j. \end{cases}$$

In matrix notation, by introducing the diagonal matrix, $X_{k-2}(x)$, whose entries are the components of vector $x^{[k-2]}$, that is, $X_{k-2}(x) \doteq \text{diag}(x^{[k-2]})$, we obtain:

$$\nabla^2 g(x) = m(m-k)\|x\|_k^{m-2k}x^{[k-1]}(x^{[k-1]})^\top + m(k-1)\|x\|_k^{m-k}X_{k-2}(x).$$

Consequently,

$$\begin{array}{rcl} \nabla^2 g(x) & = & m(m-1)X_{m-2}(x), \text{ if } m = k; \\ \nabla^2 g(x) & = & m(m-2)\|x\|_2^{m-4}xx^\top + m\|x\|_2^{m-2}I, \text{ if } k = 2. \end{array}$$

The case when C is convex.
We now assume that C is additionally convex. By Definition 3.1 and (*iii*) of Remark 3.3, any L-eigenvalue λ ensures the existence of $x \in \mathbb{R}^n$ such that (see Subsection 3.4.2)

$$\begin{cases} \mathcal{A}x^{m-1} - \lambda\|x\|_k^{m-k}\varphi_k(x) \in C^*, \\ \mathcal{A}x^m - \lambda\|x\|_k^m = 0, \\ x \in C \setminus \{0\}. \end{cases} \tag{3.53}$$

A pair (λ, x) satisfying (3.53) is termed a L-eigenpair (see Definition 3.1 in Subsection 3.4.2). In other words, $x \in C$, $x \neq 0$, is a KKT-point of (3.51) if (λ, x) is a L-eigenpar for some $\lambda \in \mathbb{R}$. According to the choice of m, k and C, such a pair takes different names, see the remarks below. The system (3.53) is a class of homogeneous complementarity problem, since:

$$\langle x, \mathcal{A}x^{m-1} - \lambda\|x\|_k^{m-k}\varphi_k(x)\rangle = 0 \Longleftrightarrow \mathcal{A}x^m - \lambda\|x\|_k^m = 0.$$

Remark 3.9. (*Z-eigenvalues/eigenvectors, $C = \mathbb{R}^n$*)
The case $k = 2$ and $C = \mathbb{R}^n$ was discussed in [122], where the term Z-eigenvalues (eigenvectors) is employed. The authors in [92] use the name l^2-eigenvalues (eigenvectors). Here, (3.53) *reduces:*

$$\begin{cases} \mathcal{A}x^{m-1} - \lambda\|x\|_2^{m-2}x = 0, \\ \mathcal{A}x^m - \lambda\|x\|_2^m = 0, \\ x \in C \setminus \{0\}, \end{cases}$$

since $\varphi_2(x) = x$ *and* $T(\mathbb{R}^n; x) = \mathbb{R}^n$.

Remark 3.10. (H*-eigenvalues/eigenvectors,* $C = \mathbb{R}^n$)
The case $k = m$ *and* $C = \mathbb{R}^n$ *was also analyzed in [122], which gives rise to* H*-eigenvalues (eigenvectors). It coincides with the notion of* l^m*-eigenvalues (eigenvectors), introduced in [92]. The homogeneous complementarity problem* (3.53) *becomes:*

$$\begin{cases} \mathcal{A}x^{m-1} - \lambda\varphi_m(x) = 0, \\ \mathcal{A}x^m - \lambda\|x\|_m^m = 0, \\ x \in C \setminus \{0\}. \end{cases}$$

Observe that when m *is even,* $\varphi_m(x) = x^{[m-1]}$, *and if* m *is odd, one gets:*

$$\varphi_m(x)_i = \begin{cases} x_i^{m-1} & \text{if } x_i \geq 0; \\ -x_i^{m-1} & \text{if } x_i < 0. \end{cases}$$

Remark 3.11. *The cases* $C = \mathbb{R}^n_+$ *and either* $k = 2$ *or* $k = m$ *were studied in [137]. In view of the additional nonnegative constraint,* $\varphi_k(x) = x^{[k-1]}$ *for all* $k \in \mathbb{N}$. *Here, problem* (3.53) *takes the form:*

$$\begin{cases} \mathcal{A}x^{m-1} - \lambda\|x\|_k^{m-k}x^{[k-1]} \geq 0, \\ \mathcal{A}x^m - \lambda\|x\|_k^m = 0, \\ x \geq 0, \ x \neq 0. \end{cases} \tag{3.54}$$

From the previous remarks and together with Proposition 3.4 and Theorem 3.1, the following result extends those appeared in [92, 122, 137], where the cases $C = \mathbb{R}^n_+$ or $C = \mathbb{R}^n$ are only considered. In what follows f and g are as in (3.52).

Proposition 3.8. *Let* $\mathcal{A}$ *be an* m*-order* n*-dimensional symmetric tensor and* C *be a convex closed cone in* $\mathbb{R}^n$. *Then*

(*a*) (*Set* $C = \mathbb{R}^n$) *For each case, the set of all* Z*-eigenvalues and that of all* H*-eigenvalues coincide with our notion of* L*-eigenvalues, as well as with that of simply eigenvalues* (*See Definition 3.1*).

(*b*) *One has, for some* $\bar{x} \in C$, $\bar{x} \neq 0$,

$$\begin{aligned} \mu_k &= \min\{\lambda \in \mathbb{R} : \lambda \text{ is } L-\text{eigenvalue}\} = \inf\{\mathcal{A}x^m - \mu_k(\|x\|_k^m - 1) : \ x \in C\} \\ &= \min\{\lambda \in \mathbb{R} : \lambda \text{ is eigenvalue}\} = \min_{\substack{x \in C \\ x \neq 0}} \frac{\mathcal{A}x^m}{\|x\|_k^m} = \frac{\mathcal{A}\bar{x}^m}{\|\bar{x}\|_k^m}. \end{aligned}$$

(*c*) $\mu_k = \max\{\lambda : \ f - \lambda g$ is copositive on $C\}$.

Remark 3.12. *Among those KKT-points for problem* (3.51), *one can identify a solution,* $\bar{x}$, *to* (3.51) *simply by using Theorem 3.5, once we impose conditions implying the pseudoconvexity of* f *and the quasiconvexiy of the function* $x \mapsto -f(\bar{x})g(x)$. *Notice that* g *is always quasiconvex, and so we must take care on the sign of* $f(\bar{x})$.

The case $C \doteq \mathbb{R}^n_+$ is of particular interest. In this situation, one can also consider the standard dual problem to (3.51):

$$\nu_S \doteq \sup_{\substack{\lambda_1 \geq 0 \\ \lambda_2 \in \mathbb{R}^n_+}} \inf_{x \in \mathbb{R}^n} \left\{ \mathcal{A}x^m + \lambda_1(\|x\|_k^m - 1) - \lambda_2^\top x \right\}.$$

We recall that, in this setting, the notions of copositivity or strict copositivity of $\mathcal{A}$ on C are referred to f on C.

Theorem 3.13. *Set $C \doteq \mathbb{R}^n_+$. Let $\mathcal{A}$ be a m-order, n-dimensional symmetric tensor, that is copositive on C but not copositive in $\mathbb{R}^n$. Then,*

$$\nu_S < 0 \leq \mu_k = \nu_k.$$

As a consequence, standard strong duality does not hold.

Proof. By assumption, $\mu_k \geq 0$ and $\mu_k = \nu_k$ (strong duality holds). We choose $\bar{x} \in \mathbb{R}^n$ such that $f(\bar{x}) < 0$. Set $\bar{\lambda}_1 \doteq -\dfrac{f(\bar{x})}{g(\bar{x})} > 0$; it follows that $f(\bar{x}) + \lambda g(\bar{x}) < 0$ for all $\lambda < \bar{\lambda}_1$ and $f(\bar{x}) + \lambda g(\bar{x}) > 0$ for all $\lambda > \bar{\lambda}_1$. Since for every $t > 0$

$$\begin{aligned} \inf_{x \in \mathbb{R}^n} \left\{ f(x) + \lambda_1 g(x) - \lambda_2^\top x) \right\} &\leq f(t\bar{x}) + \lambda_1 g(t\bar{x}) - t\lambda_2^\top \bar{x} \\ &= t^m \left[f(\bar{x}) + \lambda_1 g(\bar{x}) - \frac{1}{t^{m-1}} \lambda_2^\top \bar{x} \right], \end{aligned}$$

we obtain

$$\inf_{x \in \mathbb{R}^n} \left\{ f(x) + \lambda_1 g(x) - \lambda_2^\top x) \right\} = -\infty, \quad \forall\ \lambda_1 < \bar{\lambda}_1,\ \forall\ \lambda_2 \in \mathbb{R}^n_+.$$

Moreover,

$$\inf_{x \in \mathbb{R}^n} \left\{ f(x) + \lambda_1 (g(x) - 1) - \lambda_2^\top x) \right\} \leq -\lambda_1 \quad \forall\ \lambda_1 \geq \bar{\lambda}_1,\ \forall\ \lambda_2 \in \mathbb{R}^n_+.$$

Hence

$$\nu_S \doteq \sup_{\substack{\lambda_1 \in \mathbb{R} \\ \lambda_2 \in \mathbb{R}^n_+}} \inf_{x \in \mathbb{R}^n} \left\{ f(x) + \lambda_1 (g(x) - 1) - \lambda_2^\top x \right\} \leq -\bar{\lambda}_1 < 0 \leq \mu_k = \nu_k,$$

and the proof is complete. □

We end this section by making some comments about numerical approaches on computing eigenvalues of symmetric tensors. When only the smallest or the largest eigenvalue of a tensor is needed, one can use: the NQZ method [33] or an iterative one as proposed in [106] (for irreducible nonnegative tensors); an unconstrained optimization approach [72] for even order tensors; a SOS polynomial optimization scheme as discussed in [75] (for essentially nonnegative tensors), or the S-HOPM or SS-HOPM methods as presented in [85] and [86], respectively. For computing all the real eigenvalues of a symmetric tensor, an approach based on the Jacobian SDP relaxation method is introduced in [36].

3.8 Conclusions and future works

In this chapter an important class of positively homogeneous optimization problems is analyzed. This class of problems groups various models: specially from classical portfolio (mean-variance) problems to some of its variants appearing, when dealing with combined economic emission dispatch problems and integration of renewable energy sources. Results obtained in this chapter allow us to overcome usual convexity and regularity assumptions needed in other investigations; therefore, we can provide results, for example, when fuel cost or emission functions are not necessarily convex but given by homogeneous polynomials. Even more, as it can be seen, our results are additionally applicable for tensors eigenvalues analysis, quadratic fractional optimization problems and in general, for a class of problems defined by homogeneous functions and the geometrical constraint set given by two quadratic forms.

As future works, it would be interesting to identify and to analyze other real problems fitting in our model, specially those where more traditional approaches fail. Particularly, for the combined economic emission dispatch problem, we believe two particular subproblems must be addressed: when the quadratic emission and/or cost functions are not convex and when these functions are given by cubic polynomials instead. In the later case, is interesting to identify if a suitable homogenization formulation can be used. On another hand, the copositive re-formulation and the bisection based algorithm proposed to solve it, when $p = q$, must be enhanced. Note that, due to the constraint $g(x) = 1$, the applicability just for the cases where $p = q$ is not too restrictive and in general, a enhanced version of this algorithm could be useful for obtaining the optimal value of a problem (where not necessarily $p = q$) by solving an auxiliary problem instead. Finally, we also consider that algorithms must be studied in order to provide not just theoretical insights but solutions when other approaches fail.

Bibliography

[1] Abdo, M., Kamel, S., Ebeed, M., Yu, J., and Jurado, F. Solving Non-Smooth Optimal Power Flow Problems Using a Developed Grey Wolf Optimizer. *Energies 11*, 7 (2018).

[2] Abou El Ela, A. A., Abido, M. A., and Spea, S. R. Differential evolution algorithm for emission constrained economic power dispatch problem. *Electric Power Systems Research 80*, 10 (2010), 1286–1292.

[3] Ah King, R. T., and Rughooputh, H. C. S. Elitist multiobjective evolutionary algorithm for environmental/economic dispatch. vol. 2, pp. 1108–1114.

[4] Ah King, R. T., Rughooputh, H. C. S., and Deb, K. Evolutionary Multi-objective Environmental/Economic Dispatch: Stochastic Versus Deterministic Approaches. In *Evolutionary Multi-Criterion Optimization* (Berlin, Heidelberg, 2005), C. A. Coello Coello, A. Hernández Aguirre, and E. Zitzler, Eds., Lecture Notes in Computer Science, Springer, pp. 677–691.

[5] Ah King, R. T., Rughooputh, H. C. S., and Deb, K. Stochastic Evolutionary Multiobjective Environmental/Economic Dispatch. In *2006 IEEE International Conference on Evolutionary Computation* (2006), pp. 946–953.

[6] Ai, W., and Zhang, S. Strong Duality for the CDT Subproblem: A Necessary and Sufficient Condition. *SIAM Journal on Optimization 19*, 4 (2009), 1735–1756.

[7] AL-Kandari, A. M., and EL-Naggar, K. M. A genetic-based algorithm for optimal estimation of input–output curve parameters of thermal power plants. *Electrical Engineering 89*, 8 (2007), 585–590.

[8] Al-Khayyal, F. A. Linear, quadratic, and bilinear programming approaches to the linear complementarity problem. *European Journal of Operational Research 24*, 2 (1986), 216–227.

[9] Alrashidi, M. R., El-Naggar, K. M., and Al-Othman, A. K. Particle Swarm Optimization Based Approach for Estimating the Fuel-cost Function Parameters of Thermal Power Plants with Valve Loading Effects. *Electric Power Components and Systems 37*, 11 (2009), 1219–1230.

[10] Askarzadeh, A., and Gharibi, M. Accurate estimation of cost function parameters for thermal power plants using a novel optimization approach. *Energy Sources, Part A: Recovery, Utilization, and Environmental Effects 40*, 24 (2018), 2986–2999.

[11] Aubin, J.-P., and Frankowska, H. *Set-Valued Analysis*. Modern Birkhäuser Classics. Birkhäuser Basel, 2009.

[12] Balamurugan, R., and Subramanian, S. A Simplified Recursive Approach to Combined Economic Emission Dispatch. *Electric Power Components and Systems 36*, 1 (2007), 17–27.

[13] Bath, S. K., Dhillon, J. S., and Kothari, D. P. Fuzzy satisfying stochastic multi-objective generation scheduling by weightage pattern search methods. *Electric Power Systems Research 69*, 2 (2004), 311–320.

[14] Bath, S. K., Dhillon, J. S., and Kothari, D. P. Security Constrained Stochastic Multi-objective Optimal Power Dispatch. *International Journal of Emerging Electric Power Systems 8*, 1 (2007).

[15] Bayón, L., Grau, J. M., Ruiz, M. M., and Suárez, P. M. Optimization of SO2 and NOx Emissions in Thermal Plants. *Journal of Mathematical Chemistry 40*, 1 (2006), 29–41.

[16] Bazaraa, M. S., Sherali, H. D., and Shetty, C. M. *Nonlinear Programming: Theory and Algorithms*. John Wiley and Sons, Inc., Hoboken, NJ, USA, 2006.

[17] Beck, A. On the convexity of a class of quadratic mappings and its application to the problem of finding the smallest ball enclosing a given intersection of balls. *Journal of Global Optimization 39*, 1 (2007), 113–126.

[18] Best, M. J., and Ding, B. Global and Local Quadratic Minimization. *Journal of Global Optimization 10*, 1 (1997), 77–90.

[19] Bhattacharya, A., and Kojima, S. Power sector investment risk and renewable energy: A Japanese case study using portfolio risk optimization method. *Energy Policy 40* (2012), 69–80.

[20] Björck, A. *Numerical Methods for Least Squares Problems*. Society for Industrial and Applied Mathematics, 1996.

[21] Bomze, I. M. Global Escape Strategies for Maximizing Quadratic Forms over a Simplex. *Journal of Global Optimization 11*, 3 (1997), 325–338.

[22] Bomze, I. M. On Standard Quadratic Optimization Problems. *Journal of Global Optimization 13*, 4 (1998), 369–387.

[23] Bomze, I. M. Copositive Relaxation Beats Lagrangian Dual Bounds in Quadratically and Linearly Constrained Quadratic Optimization Problems. *SIAM Journal on Optimization 25*, 3 (2015), 1249–1275.

[24] Bomze, I. M. Copositivity for second-order optimality conditions in general smooth optimization problems. *Optimization 65*, 4 (2016), 779–795.

[25] Bomze, I. M., and De Klerk, E. Solving Standard Quadratic Optimization Problems via Linear, Semidefinite and Copositive Programming. *Journal of Global Optimization 24*, 2 (2002), 163–185.

[26] Bomze, I. M., Dür, M., de Klerk, E., Roos, C., Quist, A. J., and Terlaky, T. On Copositive Programming and Standard Quadratic Optimization Problems. *Journal of Global Optimization 18*, 4 (2000), 301–320.

[27] Bomze, I. M., Locatelli, M., and Tardella, F. New and old bounds for standard quadratic optimization: dominance, equivalence and incomparability. *Mathematical Programming 115*, 1 (2008), 31.

[28] Cárcamo, G., and Flores-Bazán, F. Strong duality and KKT conditions in nonconvex optimization with a single equality constraint and geometric constraint. *Mathematical Programming 168*, 1 (2018), 369–400.

[29] Carrillo-Galvez, A., and Flores-Bazán, F. Non convex homogeneous optimization: a general framework and applications to quadratic fractional programming and tensor eigenvalues analysis. *Preprint 2021-14, Departamento de Ingeniería Matemática, Universidad de Concepción, Chile* (2021), Available at: https://www.ing–mat.udec.cl.

[30] Carrillo-Galvez, A., Flores-Bazán, F., and López, E. A duality theory approach to the environmental/economic dispatch problem. *Electric Power Systems Research 184* (2020), 106285.

[31] Carrillo-Galvez, A., Flores-Bazán, F., and López, E. An analytical approach to the environmental/economic dispatch problem. In *2019 IEEE CHILEAN Conference on Electrical, Electronics Engineering, Information and Communication Technologies (CHILECON)* (2019), pp. 1–5.

[32] Carrillo-Galvez, A., Flores-Bazán, F., and López, E. On the solution of the environmental/economic dispatch problem using lagrangian duality. In *2020 IEEE International Conference on Industrial Technology (ICIT)* (2020), pp. 619–623.

[33] Chang, K.-C., Pearson, K. J., and Zhang, T. Primitivity, the Convergence of the NQZ Method, and the Largest Eigenvalue for Nonnegative Tensors. *SIAM Journal on Matrix Analysis and Applications* (2011).

[34] Chen, H. Y. K., and Postel, C. E. On-Line Parameter Identification of Input Output Curves for Thermal Units. *IEEE Transactions on Power Systems 1*, 2 (1986), 221–224.

[35] Cottle, R. W., and Dantzig, G. B. Complementary pivot theory of mathematical programming. *Linear Algebra and its Applications 1*, 1 (1968), 103–125.

[36] Cui, C.-F., Dai, Y.-H., and Nie, J. All Real Eigenvalues of Symmetric Tensors. *SIAM Journal on Matrix Analysis and Applications 35*, 4 (2014), 1582–1601.

[37] Cui, Y., Geng, Z., Zhu, Q., and Han, Y. Review: Multi-objective optimization methods and application in energy saving. *Energy 125* (2017), 681–704.

[38] Deb, K., Pratap, A., Agarwal, S., and Meyarivan, T. A fast and elitist multiobjective genetic algorithm: NSGA-II. *IEEE Transactions on Evolutionary Computation 6*, 2 (2002), 182–197.

[39] Dei Rossi, J. Prediction intervals for summed totals. Tech. rep., RAND Corp. Santa Monica, CA, 1968.

[40] Dhillon, J. S., Parti, S. C., and Kothari, D. P. Stochastic economic emission load dispatch. *Electric Power Systems Research 26*, 3 (1993), 179–186.

[41] Dhillon, J. S., Parti, S. C., and Kothari, D. P. Multiobjective Decision Making in Stochastic Economic Dispatch. *Electric Machines and Power Systems 23*, 3 (1995), 289–301.

[42] Di, S., and Poliquin, R. Contingent cone to a set defined by equality and inequality constraints at a Fréchet differentiable point. *Journal of Optimization Theory and Applications 81*, 3 (1994), 469–478.

[43] Durai, S., Subramanian, S., and Ganesan, S. Improved parameters for economic dispatch problems by teaching learning optimization. *International Journal of Electrical Power and Energy Systems 67* (2015), 11–24.

[44] Durai, S., Subramanian, S., and Ganesan, S. Preferred Economic Dispatch of Thermal Power Units. *Journal of Power and Energy Engineering 3*, 11 (2015), 47–69.

[45] Durillo, J. J., Zhang, Y., Alba, E., Harman, M., and Nebro, A. J. A study of the bi-objective next release problem. *Empirical Software Engineering 16*, 1 (2011), 29–60.

[46] Ehrgott, M. *Multicriteria Optimization*, 2nd ed. Springer, 2005.

[47] El-hawary, M. E., and Mansour, S. Y. Performance Evaluation of Parameter Estimation Algorithms for Economic Operation of Power Systems. *IEEE Transactions on Power Apparatus and Systems PAS-101*, 3 (1982), 574–582.

[48] EL-Naggar, K. M., AlRashidi, M. R., and Al-Othman, A. K. Estimating the input–output parameters of thermal power plants using PSO. *Energy Conversion and Management 50*, 7 (2009), 1767–1772.

[49] El-Shibini, M., and Osman, Z. H. A novel technique to estimate the fuel cost functions for economic operation of power systems. *International Journal of Electrical Power and Energy Systems 11*, 2 (1989), 109–114.

[50] Elaiw, A. M., Xia, X., and Shehata, A. M. Application of model predictive control to optimal dynamic dispatch of generation with emission limitations. *Electric Power Systems Research 84*, 1 (2012), 31–44.

[51] Elton, J. H. Indefinite Quadratic Forms and the Invariance of the Interval in Special Relativity. *The American Mathematical Monthly 117*, 6 (2010), 540–547.

[52] Fallahi, S., and Salahi, M. On the Indefinite Quadratic Fractional Optimization with Two Quadratic Constraints. *Journal of Optimization Theory and Applications 162*, 1 (2014), 249–256.

[53] Fisher, R. A., and Cornish, E. A. The Percentile Points of Distributions Having Known Cumulants. *Technometrics 2*, 2 (1960), 209–225.

[54] Flores-Bazán, F., Cárcamo, G., and Caro, S. Extensions of the Standard Quadratic Optimization Problem: Strong Duality, Optimality, Hidden Convexity and S-lemma. *Applied Mathematics and Optimization 81*, 2 (2020), 383–408.

[55] Flores-Bazán, F., Echegaray, W., Flores-Bazán, F., and Ocaña, E. Primal or dual strong-duality in nonconvex optimization and a class of quasiconvex problems having zero duality gap. *Journal of Global Optimization 69*, 4 (2017), 823–845.

[56] Flores-Bazán, F., and López, R. The Linear Complementarity Problem Under Asymptotic Analysis. *Mathematics of Operations Research 30*, 1 (2005), 73–90.

[57] Flores-Bazán, F., and Mastroeni, G. Characterizing FJ and KKT Conditions in Nonconvex Mathematical Programming with Applications. *SIAM Journal on Optimization 25*, 1 (2015), 647–676.

[58] Flores-Bazán, F., Mastroeni, G., and Vera, C. Proper or Weak Efficiency via Saddle Point Conditions in Cone-Constrained Nonconvex Vector Optimization Problems. *Journal of Optimization Theory and Applications 181*, 3 (2019), 787–816.

[59] Flores-Bazán, F., and Opazo, F. Characterizing the convexity of joint-range for a pair of inhomogeneous quadratic functions and strong duality. *Minimax Theory and its Applications 1* (2016), 257–290.

[60] Flores-Bazán, F., and Opazo, F. Characterizing convexity of images for quadratic-linear mappings with applications in nonconvex quadratic optimization. *SIAM Journal on Optimization 1* (2021), In press.

[61] Gay, D. M. Computing Optimal Locally Constrained Steps. *SIAM Journal on Scientific and Statistical Computing 2*, 2 (1981), 186–197.

[62] Geng, Z., Chen, Q., Xia, Q., Kirschen, D. S., and Kang, C. Environmental Generation Scheduling Considering Air Pollution Control Technologies and Weather Effects. *IEEE Transactions on Power Systems 32*, 1 (2017), 127–136.

[63] Geng, Z., Conejo, A. J., Chen, Q., and Kang, C. Power generation scheduling considering stochastic emission limits. *International Journal of Electrical Power and Energy Systems 95* (2018), 374–383.

[64] Gent, M. R., and Lamont, J. W. Minimum-Emission Dispatch. *IEEE Transactions on Power Apparatus and Systems PAS-90*, 6 (1971), 2650–2660.

[65] Ghasemi, M., Aghaei, J., Akbari, E., Ghavidel, S., and Li, L. A differential evolution particle swarm optimizer for various types of multi-area economic dispatch problems. *Energy 107* (2016), 182–195.

[66] Gibbons, L. E., Hearn, D. W., Pardalos, P. M., and Ramana, M. V. Continuous Characterizations of the Maximum Clique Problem. *Mathematics of Operations Research 22*, 3 (1997), 754–768.

[67] Glimn, A. F., Kirchmayer, L. K., Peterson, V. R., and Stagg, G. W. Accuracy Considerations in Economic Dispatching of Power Systems - I [includes discussion]. *Transactions of the American Institute of Electrical Engineers. Part III: Power Apparatus and Systems 75*, 3 (1956), 1125–1137.

[68] Guttikunda, S. K., and Jawahar, P. Atmospheric emissions and pollution from the coal-fired thermal power plants in India. *Atmospheric Environment 92* (2014), 449–460.

[69] Güven c, U., Sönmez, Y., Duman, S., and Yörükeren, N. Combined economic and emission dispatch solution using gravitational search algorithm. *Scientia Iranica 19*, 6 (2012), 1754–1762.

[70] Güvenç, U. Combined economic emission dispatch solution using genetic algorithm based on similarity crossover. *Scientific Research and Essays 5*, 17 (2010), 2451–2456.

[71] Hamdy, M., Nguyen, A.-T., and Hensen, J. L. A performance comparison of multi-objective optimization algorithms for solving nearly-zero-energy-building design problems. *Energy and Buildings 121* (2016), 57–71.

[72] Han, L. An unconstrained optimization approach for finding real eigenvalues of even order symmetric tensors. *Numerical Algebra, Control and Optimization 3*, 3 (2013), 583.

[73] He, Z., Zhu, P.-F., and Park, S.-H. A robust desirability function method for multi-response surface optimization considering model uncertainty. *European Journal of Operational Research 221*, 1 (2012), 241–247.

[74] Hiriart-Urruty, J.-B., and Seeger, A. A Variational Approach to Copositive Matrices. *SIAM Review 52*, 4 (2010), 593–629.

[75] Hu, S., Li, G., Qi, L., and Song, Y. Finding the Maximum Eigenvalue of Essentially Nonnegative Symmetric Tensors via Sum of Squares Programming. *Journal of Optimization Theory and Applications 158*, 3 (2013), 717–738.

[76] Hu, S., Sun, H., Peng, F., Zhou, W., Cao, W., Su, A., Chen, X., and Sun, M. Optimization Strategy for Economic Power Dispatch Utilizing Retired EV Batteries as Flexible Loads. *Energies 11*, 7 (2018).

[77] Ibaraki, T., and Katoh, N. *Resource allocation problems: algorithmic approaches*. MIT Press, Cambridge, MA, USA, 1988.

[78] IEEE Committee. Present practices in the economic operation of power systems. Report Paper 71TP 97-PWR, IEEE Winter Power Meeting, New York, 1971.

[79] Ioannou, A., Angus, A., and Brennan, F. Risk-based methods for sustainable energy system planning: A review. *Renewable and Sustainable Energy Reviews 74* (2017), 602–615.

[80] Jadhav, H., and Roy, R. Gbest guided artificial bee colony algorithm for environmental/economic dispatch considering wind power. *Expert Systems with Applications 40*, 16 (2013), 6385–6399.

[81] Jeyakumar, D., Jayabarathi, T., and Raghunathan, T. Particle swarm optimization for various types of economic dispatch problems. *International Journal of Electrical Power and Energy Systems 28*, 1 (2006), 36–42.

[82] Jeyakumar, V., and Li, G. Y. Trust-region problems with linear inequality constraints: exact SDP relaxation, global optimality and robust optimization. *Mathematical Programming 147*, 1 (2014), 171–206.

[83] Khaled Khamees, A., Badra, N. M., and Abdelaziz, A. Y. Optimal Power Flow Methods: A Comprehensive Survey. *International Electrical Engineering Journal 7*, 4 (2016), 2228–2239.

[84] Koehler, K. J. A Simple Approximation for the Percentiles of the t Distribution. *Technometrics 25*, 1 (1983), 103–105.

[85] Kofidis, E., and Regalia, P. A. On the Best Rank-1 Approximation of Higher-Order Supersymmetric Tensors. *SIAM Journal on Matrix Analysis and Applications* (2006).

[86] Kolda, T. G., and Mayo, J. R. Shifted Power Method for Computing Tensor Eigenpairs. *SIAM Journal on Matrix Analysis and Applications 32*, 4 (2011), 1095–1124.

[87] Landsman, Z. Minimization of the root of a quadratic functional under a system of affine equality constraints with application to portfolio management. *Journal of Computational and Applied Mathematics 220*, 1 (2008), 739–748.

[88] Landsman, Z., and Makov, U. Translation-invariant and positive-homogeneous risk measures and optimal portfolio management. *The European Journal of Finance 17*, 4 (2011), 307–320.

[89] Lasserre, J. B. Global Optimization with Polynomials and the Problem of Moments. *SIAM Journal on Optimization 11*, 3 (2001), 796–817.

[90] Liang, Z.-X., and Glover, J. Improved cost functions for economic dispatch compensations (on power systems). *IEEE Transactions on Power Systems 6*, 2 (1991), 821–829.

[91] Liao, G.-C. Solve environmental economic dispatch of Smart MicroGrid containing distributed generation system – Using chaotic quantum genetic algorithm. *International Journal of Electrical Power and Energy Systems 43*, 1 (2012), 779–787.

[92] Lim, L.-H. Singular values and eigenvalues of tensors: a variational approach. In *1st IEEE International Workshop on Computational Advances in Multi-Sensor Adaptive Processing, 2005.* (2005), pp. 129–132.

[93] Liu, W., Wynn, H. P., and Hayter, A. J. Statistical inferences for linear regression models when the covariates have functional relationships: polynomial regression. *Journal of Statistical Computation and Simulation 78*, 4 (2008), 315–324.

[94] Mahdad, B., Srairi, K., and Bouktir, T. Optimal power flow for large-scale power system with shunt FACTS using efficient parallel GA. *International Journal of Electrical Power and Energy Systems 32*, 5 (2010), 507–517.

[95] Mahdi, F. P., Vasant, P., Abdullah-Al-Wadud, M., Watada, J., and Kallimani, V. A quantum-inspired particle swarm optimization approach for environmental/economic power dispatch problem using cubic criterion function. *International Transactions on Electrical Energy Systems 28*, 3 (2018), e2497.

[96] Mahdi, F. P., Vasant, P., Kallimani, V., Watada, J., Fai, P. Y. S., and Abdullah-Al-Wadud, M. A holistic review on optimization strategies for combined economic emission dispatch problem. *Renewable and Sustainable Energy Reviews 81* (2018), 3006–3020.

[97] Markowitz, H. Portfolio Selection. *The Journal of Finance 7*, 1 (1952), 77–91.

[98] Markowitz, H. The general mean-variance portfolio selection problem. In *Mathematical Models in Finance*. Chapman and Hall, 1995.

[99] Matskani, E., Sidiropoulos, N. D., Luo, Z.-q., and Tassiulas, L. Convex approximation techniques for joint multiuser downlink beamforming and admission control. *IEEE Transactions on Wireless Communications 7*, 7 (2008), 2682–2693.

[100] Miller, B. G. The Effect of Coal Usage on Human Health and the Environment. In *Clean Coal Engineering Technology (Second Edition)*, B. G. Miller, Ed. Butterworth-Heinemann, 2017, pp. 105–144.

[101] Modiri-Delshad, M., and Rahim, N. A. Multi-objective backtracking search algorithm for economic emission dispatch problem. *Applied Soft Computing 40* (2016), 479–494.

[102] Mohamed, F. A., and Koivo, H. N. System modelling and online optimal management of MicroGrid using Mesh Adaptive Direct Search. *International Journal of Electrical Power and Energy Systems 32*, 5 (2010), 398–407.

[103] Mohammadian, H., Rahimi, A., and Seyyed, M. M. Solving Environmental/Economic Power Dispatch Problem by a Trust Region Based Augmented Lagrangian Method. *Iranian Journal of Electrical and Electronic Engineering 8*, 2 (2012), 176–187.

[104] Montgomery, D. C., Peck, E. A., and Vining, G. *Introduction to Linear Regression Analysis*, 5th ed. Wiley Series in Probability and Statistic. Wiley, New Jersey, 2012.

[105] Motzkin, T. S., and Straus, E. G. Maxima for Graphs and a New Proof of a Theorem of Turán. *Canadian Journal of Mathematics 17* (1965), 533–540.

[106] Ng, M., Qi, L., and Zhou, G. Finding the Largest Eigenvalue of a Nonnegative Tensor. *SIAM Journal on Matrix Analysis and Applications 31*, 3 (2009), 1090–1099.

[107] Nguyen, V.-B., Nguyen, T. N., and Sheu, R.-L. Strong duality in minimizing a quadratic form subject to two homogeneous quadratic inequalities over the unit sphere. *Journal of Global Optimization 76*, 1 (2020), 121–135.

[108] Nie, J. Sum of squares methods for minimizing polynomial forms over spheres and hypersurfaces. *Frontiers of Mathematics in China 7*, 2 (2012), 321–346.

[109] Noakes, F., and Arismunandar, A. Bibliography on Optimum Operation of Power Systems: 1919-1959. *Transactions of the American Institute of Electrical Engineers. Part III: Power Apparatus and Systems 81*, 3 (1962), 864–871.

[110] Olson, D. L., Wei, J. C., and White, E. M. Prediction-Interval Constrained Programming. *Journal of Quality Technology 20*, 2 (1988), 90–97.

[111] Olusegun, E., and Sivaram, M. A modification of the Fisher-Cornish approximation for the student t percentiles. *Communications in Statistics-Simulation and Computation 16*, 4 (1987), 1123–1132.

[112] Orero, S. O., and Irving, M. R. Economic dispatch of generators with prohibited operating zones: a genetic algorithm approach. *IEE Proceedings - Generation, Transmission and Distribution 143*, 6 (1996), 529–534.

[113] Panigrahi, B., Pandi, V. R., Sharma, R., Das, S., and Das, S. Multiobjective bacteria foraging algorithm for electrical load dispatch problem. *Energy Conversion and Management 52*, 2 (2011), 1334–1342.

[114] Pardalos, P. M., Ye, Y., and Han, C. G. Algorithms for the solution of quadratic knapsack problems. *Linear Algebra and Its Applications 152*, C (1991), 69–91.

[115] Peng, J.-M., and Yuan, Y.-X. Optimality Conditions for the Minimization of a Quadratic with Two Quadratic Constraints. *SIAM Journal on Optimization 7*, 3 (1997), 579–594.

[116] Pérez, R., Watts, D., and Flores, Y. Planning in a changing environment: Applications of portfolio optimisation to deal with risk in the electricity sector. *Renewable and Sustainable Energy Reviews 82* (2018), 3808–3823.

[117] Pérez, R., Watts, D., and Negrete-Pincetic, M. Portfolio applications in electricity markets review: Private investor and manager perspective trends. *Renewable and Sustainable Energy Reviews 81* (2018), 192–204.

[118] Petrovic, R., and Kralj, B. Economic and environmental power dispatch. *European Journal of Operational Research 64*, 1 (1993), 2–11.

[119] Pinsler, P. Über das Vorkommen definiter und semidefiniter Formen in Scharen quadratischer Formen. *Commentarii Mathematici Helvetici 9*, 1 (1936), 188–192.

[120] Pólik, I., and Terlaky, T. A Survey of the S-Lemma. *SIAM Review 49*, 3 (2007), 371–418.

[121] Preisig, J. C. Copositivity and the Minimization of Quadratic Functions with Nonnegativity and Quadratic Equality Constraints. *SIAM Journal on Control and Optimization 34*, 4 (1996), 1135–1150.

[122] Qi, L. Eigenvalues of a real supersymmetric tensor. *Journal of Symbolic Computation 40*, 6 (2005), 1302–1324.

[123] Qu, B., Zhu, Y., Jiao, Y., Wu, M., Suganthan, P., and Liang, J. A survey on multi-objective evolutionary algorithms for the solution of the environmental/economic dispatch problems. *Swarm and Evolutionary Computation 38* (2018), 1–11.

[124] Rockafellar, R. T. *Conjugate Duality and Optimization*. CBMS-NSF Regional Conference Series in Applied Mathematics. Society for Industrial and Applied Mathematics, 1974.

[125] Sayah, S., and Hamouda, A. Novel application of differential evolution algorithm for estimating fuel cost function of thermal generating units. In *2015 Third World Conference on Complex Systems (WCCS)* (2015), pp. 1–6.

[126] Sayah, S., and Hamouda, A. Efficient method for estimation of smooth and nonsmooth fuel cost curves for thermal power plants. *International Transactions on Electrical Energy Systems 28*, 3 (2018), e2498.

[127] Schlüter, S., and Fischer, M. A Tail Quantile Approximation for the Student t Distribution. *Communications in Statistics-Theory and Methods 41*, 15 (2012), 2617–2625.

[128] Secui, D., Hora, C., Bendea, G., and Bendea, C. Parameter estimation using a modified whale optimization algorithm for input-output curves of thermal and hydro power plants. *International Transactions on Electrical Energy Systems 30*, 2 (2020), e12188.

[129] Seeger, A. Eigenvalue analysis of equilibrium processes defined by linear complementarity conditions. *Linear Algebra and its Applications 292*, 1 (1999), 1–14.

[130] Seeger, A., and Torki, M. Local minima of quadratic forms on convex cones. *Journal of Global Optimization 44*, 1 (2009), 1–28.

[131] Sidiropoulos, N., Davidson, T., and Luo, Z.-Q. Transmit beamforming for physical-layer multicasting. *IEEE Transactions on Signal Processing 54*, 6 (2006), 2239–2251.

[132] Silva, C., and Quiroz, A. Optimization of the atmospheric pollution monitoring network at Santiago de Chile. *Atmospheric Environment 37*, 17 (2003), 2337–2345.

[133] Sivasubramani, S., and Swarup, K. Environmental/economic dispatch using multi-objective harmony search algorithm. *Electric Power Systems Research 81*, 9 (2011), 1778–1785.

[134] Smoktunowicz, A., and Wróbel, I. Numerical aspects of computing the Moore-Penrose inverse of full column rank matrices. *BIT Numerical Mathematics 52*, 2 (2012), 503–524.

[135] Soliman, S., and Al-Kandari, A. Kalman filtering algorithm for on-line identification of input-output curves for thermal power plant. In *Proceedings of 8th Mediterranean Electrotechnical Conference on Industrial Applications in Power Systems, Computer Science and Telecommunications (MELECON 96)* (1996), vol. 3, pp. 1588–1593 vol.3.

[136] Soliman, S. A., Emam, S. E. A., and Christensen, G. S. Optimization of the optimal coefficients of non-monotonically increasing incremental cost curves. *Electric Power Systems Research 21*, 2 (1991), 99–106.

[137] Song, Y., and Qi, L. Eigenvalue analysis of constrained minimization problem for homogeneous polynomial. *Journal of Global Optimization 64*, 3 (2016), 563–575.

[138] Song, Y. H., and Yu, I.-K. Dynamic load dispatch with voltage security and environmental constraints. *Electric Power Systems Research 43*, 1 (1997), 53–60.

[139] Sönmez, Y. Estimation of fuel cost curve parameters for thermal power plants using the ABC algorithm. *Turkish Journal of Electrical Engineering and Computer Sciences 21*, Sup.1 (2013), 1827–1841.

[140] Sönmez, Y., and Unal, M. Estimation of Smooth and Non-smooth Fuel Cost Function Parameters Using Improved Symbiotic Organisms Search Algorithm. *Journal of Electrical Engineering and Technology 15*, 1 (2020), 13–25.

[141] Sorensen, D. C. Newton's Method with a Model Trust Region Modification. *SIAM Journal on Numerical Analysis 19*, 2 (1982), 409–426.

[142] Stanzani, A., Balbo, A., Nepomuceno, L., and Baptista, E. Solving the Multiobjective Environmental/Economic Dispatch Problem using Weighted Sum and e -Constraint Strategies and a Predictor-Corrector Primal-Dual Interior Point Method.

[143] Talaq, J. H., El-Hawary, F., and El-Hawary, M. E. A summary of environmental/economic dispatch algorithms. *IEEE Transactions on Power Systems 9*, 3 (1994), 1508–1516.

[144] Tang, L., Qu, J., Mi, Z., Bo, X., Chang, X., Anadon, L. D., Wang, S., Xue, X., Li, S., Wang, X., and Zhao, X. Substantial emission reductions from Chinese power plants after the introduction of ultra-low emissions standards. *Nature Energy 4*, 11 (2019), 929–938.

[145] Tihansky, D. P. Confidence Assessment of Military Airframe Cost Predictions. *Operations Research* (1976).

[146] Vaisakh, K., Praveena, P., Rama, S., and Meah, K. Solving dynamic economic dispatch problem with security constraints using bacterial foraging PSO-DE algorithm. *International Journal of Electrical Power and Energy Systems 39*, 1 (2012), 56–67.

[147] Vanithasri, M., Balamurugan, R., and Lakshminarasimman, L. Modified radial movement optimization (MRMO) technique for estimating the parameters of fuel cost function in thermal power plants. *Engineering Science and Technology, an International Journal 19*, 4 (2016), 2035–2042.

[148] Vemuri, S., and Hill, E. Sensitivity analysis of optimum operation of hydro-thermal plants. *IEEE Transactions on Power Apparatus and Systems 96*, 2 (1977), 688–696.

[149] Wang, L., and Singh, C. Stochastic economic emission load dispatch through a modified particle swarm optimization algorithm. *Electric Power Systems Research 78*, 8 (2008), 1466–1476.

[150] Wei, J. C., Olson, D. L., and White, E. M. Simultaneous Optimization in Process Quality Control via Prediction-interval Constrained Programming. *Journal of the Operational Research Society 41*, 12 (1990), 1161–1167.

[151] Welch, B. L. The Significance of the Difference Between Two Means when the Population Variances are Unequal. *Biometrika 29*, 3/4 (1938), 350–362.

[152] Welch, B. L. The Generalization of Student's Problem when Several Different Population Variances are Involved. *Biometrika 34*, 1/2 (1947), 28–35.

[153] Wiebking, R. Economic dispatch of thermal power systems using geometric programming. *European Journal of Operational Research 2*, 3 (1978), 168–171.

[154] Wood, A. J., Wollenberg, B. F., and Sheblé, G. B. *Power Generation, Operation, and Control*. John Wiley and Sons, 2013.

[155] Xia, X., and Elaiw, A. Optimal dynamic economic dispatch of generation: A review. *Electric Power Systems Research 80*, 8 (2010), 975–986.

[156] Xia, Y., Ghiocel, S. G., Dotta, D., Shawhan, D., Kindle, A., and Chow, J. H. A Simultaneous Perturbation Approach for Solving Economic Dispatch Problems With Emission, Storage, and Network Constraints. *IEEE Transactions on Smart Grid 4*, 4 (2013), 2356–2363.

[157] Xia, Y., Wang, S., and Sheu, R.-L. S-lemma with equality and its applications. *Mathematical Programming 156*, 1 (2016), 513–547.

[158] Xu, D., and Albin, S. L. Robust Optimization of Experimentally Derived Objective Functions. *IIE Transactions 35*, 9 (2003), 793–802.

[159] Yadav, S., and Prakash, R. Status and Environmental Impact of Emissions from Thermal Power Plants in India. *Environmental Forensics 15*, 3 (2014), 219–224.

[160] Yakubovich, V. A. S-procedure in nonlinear control theory. *Vestnik Leningrad. Univ. 1* (1971), 62–77.

[161] Yakubovich, V. A. S-procedure in nonlinear control theory. *Vestnik Leningrad. Univ. 1* (1977), 73–93.

[162] Yuan, Y.-X. Recent advances in trust region algorithms. *Mathematical Programming 151*, 1 (2015), 249–281.

[163] Zhou, B., Chan, K. W., Yu, T., and Chung, C. Y. Equilibrium-Inspired Multiple Group Search Optimization With Synergistic Learning for Multiobjective Electric Power Dispatch. *IEEE Transactions on Power Systems 28*, 4 (2013), 3534–3545.

[164] Zhou, J., Wang, C., Li, Y., Wang, P., Li, C., Lu, P., and Mo, L. A multi-objective multi-population ant colony optimization for economic emission dispatch considering power system security. *Applied Mathematical Modelling 45* (2017), 684–704.

[165] Zou, D., Li, S., Li, Z., and Kong, X. A new global particle swarm optimization for the economic emission dispatch with or without transmission losses. *Energy Conversion and Management 139* (2017), 45–70.