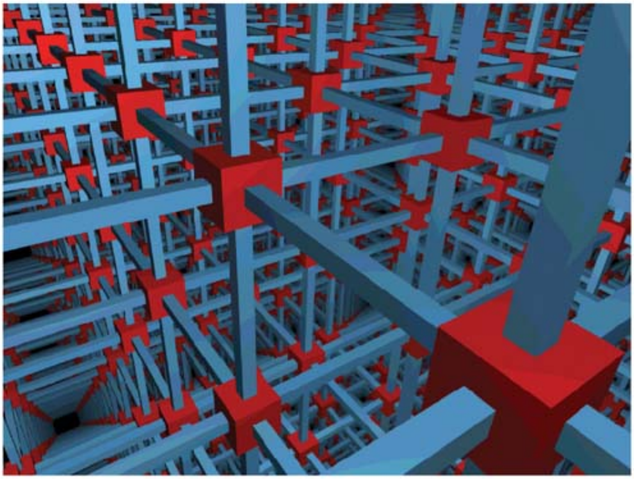
Princeton Series in APPLIED MATHEMATICS

Robust Optimization



Aharon Ben-Tal
Laurent El Ghaoui
Arkadi Nemirovski

Robust Optimization

Princeton Series in Applied Mathematics

Series Editors: Ingrid Daubechies (Princeton University); Weinan E (Princeton University); Jan Karel Lenstra (Eindhoven University); Endre Süli (University of Oxford)

The Princeton Series in Applied Mathematics publishes high quality advanced texts and monographs in all areas of applied mathematics. Books include those of a theoretical and general nature as well as those dealing with the mathematics of specific applications areas and real-world situations.

Chaotic Transitions in Deterministic and Stochastic Dynamical Systems: Applications of Melnikov Processes in Engineering, Physics, and Neuroscience, Emil Simiu

Self-Regularity: A New Paradigm for Primal-Dual Interior Point Algorithms, Jiming Peng, Cornelis Roos, and Tamas Terlaky

Selfsimilar Processes, Paul Embrechts and Makoto Maejima

Analytic Theory of Global Bifurcation: An Introduction, Boris Buffoni and John Toland

Entropy, Andreas Greven, Gerhard Keller, and Gerald Warnecke, editors

Auxiliary Signal Design for Failure Detection, Stephen L. Campbell and Ramine Nikoukhah

Max Plus at Work Modeling and Analysis of Synchronized Systems: A Course on Max-Plus Algebra and Its Applications, Bernd Heidergott, Geert Jan Olsder, and Jacob van der Woude

Optimization: Insights and Applications, Jan Brinkhuis and Vladimir Tikhomirov

Thermodynamics: A Dynamical Systems Approach, Wassim M. Haddad, VijaySekhar Chellaboina, and Sergey G. Nersesov

Impulsive and Hybrid Dynamical Systems Stability, Dissipativity, and Control, Wassim M. Haddad, VijaySekhar Chellaboina, and Sergey G. Nersesov

Distributed Control of Robotic Networks: A Mathematical Approach to Motion Coordination Algorithms, Francesco Bullo, Jorge Cortés, and Sonia Martínez

Genomic Signal Processing, Ilya Shmulevich and Edward Dougherty

Positive Definite Matrices, Rajendra Bhatia

The Traveling Salesman Problem: A Computational Study, David L. Applegate, Robert E. Bixby, Vasek Chvatal, and William J. Cook

Wave Scattering by Time-Dependent Perturbations: An Introduction, G. F. Roach

Algebraic Curves over a Finite Field, J.W.P. Hirschfeld, G. Korchmáros, and F. Torres

Distributed Control of Robotic Networks: A Mathematical Approach to Motion Coordination Algorithms, Francesco Bullo, Jorge Cortés, and Sonia Martínez

Robust Optimization, Aharon Ben-Tal, Laurent El Ghaoui, and Arkadi Nemirovski

Robust Optimization

Aharon Ben-Tal

Laurent El Ghaoui

Arkadi Nemirovski

Princeton University Press

Princeton and Oxford

Published by Princeton University Press, 41 William Street, Princeton, New Jersey

08540

In the United Kingdom: Princeton University Press, 6 Oxford Street, Woodstock,

Oxfordshire OX20 1TW

Library of Congress Cataloging-in-Publication Data

Ben-Tal, A..
Robust optimization / Aharon Ben-Tal, Laurent El Ghaoui, Arkadi Nemirovski.
p. cm.
ISBN 978-0-691-14368-2 (hardcover : alk. paper) 1. Robust optimization. 2. Linear programming. I. El Ghaoui, Laurent. II. Nemirovskii, Arkadii Semenovich. III. Title.
QA402.5.B445 2009
519.6--dc22
2009013229

The publishers would like to acknowledge the authors of this volume for providing the

electronic files from which this book was printed

Printed on acid-free paper. ∞

press.princeton.edu

Printed in the United States of America

10 9 8 7 6 5 4 3 2 1

Contents

Preface

To be uncertain is to be uncomfortable,
but to be certain is to be ridiculous.
Chinese proverb

This book is devoted to *Robust Optimization* — a specific and relatively novel methodology for handling optimization problems with *uncertain data*. The primary goal of this Preface is to provide the reader with a first impression of what the story is about:

• what is the phenomenon of data uncertainty and why it deserves a dedicated treatment,

• how this phenomenon is treated in Robust Optimization, and how this treatment compares to those offered by more traditional methodologies for handling data uncertainty.

The secondary, quite standard, goal is to outline the main topics of the book and describe its contents.

A. Data Uncertainty in Optimization

The very first question we intend to address here is whether the underlying phenomenon — data uncertainty — is worthy of special treatment. To answer this question, consider a simple example — problem `PILOT4` from the well-known `NETLIB` library. This is a Linear Programming problem with 1,000 variables and 410 constraints; one of the constraints (# 372) is:

$$\begin{array}{l} a^T x \equiv -15.79081x_{826} - 8.598819x_{827} - 1.88789x_{828} - 1.362417x_{829} - 1.526049x_{830} \\ -0.031883x_{849} - 28.725555x_{850} - 10.792065x_{851} - 0.19004x_{852} - 2.757176x_{853} \\ -12.290832x_{854} + 717.562256x_{855} - 0.057865x_{856} - 3.785417x_{857} - 78.30661x_{858} \\ -122.163055x_{859} - 6.46609x_{860} - 0.48371x_{861} - 0.615264x_{862} - 1.353783x_{863} \\ -84.644257x_{864} - 122.459045x_{865} - 43.15593x_{866} - 1.712592x_{870} - 0.401597x_{871} \\ +x_{880} - 0.946049x_{898} - 0.946049x_{916} \geq b \equiv 23.387405. \end{array} \tag{C}$$

The related *nonzero* coordinates of the optimal solution x^* of the problem, as reported by `CPLEX`, are as follows:

$$\begin{array}{lll} x^*_{826} = 255.6112787181108 & x^*_{827} = 6240.488912232100 & x^*_{828} = 3624.613324098961 \\ x^*_{829} = 18.20205065283259 & x^*_{849} = 174397.0389573037 & x^*_{870} = 14250.00176680900 \\ x^*_{871} = 25910.00731692178 & x^*_{880} = 104958.3199274139. & \end{array}$$

Note that within machine precision x^* makes (C) an equality.

Observe that most of the coefficients in (C) are "ugly reals" like -15.79081 or -84.644257. Coefficients of this type usually (and `PILOT4` is not an exception) characterize certain technological devices/processes, forecasts for future demands, etc., and as such *they could hardly be known to high accuracy*. It is quite natural to assume that the "ugly coefficients" are in fact *uncertain* — they coincide with the "true" values of the corresponding data within accuracy of 3 to 4 digits, not more. The only exception is the coefficient `1` of $\mathtt{x}_{880}$; it perhaps reflects the structure of the problem and is therefore exact, that is certain.

Assuming that the uncertain entries of a are, say, 0.1%-accurate approximations of unknown entries of the "true" vector of coefficients $\widetilde{a}$, let us look what would be the effect of this uncertainty on the validity of the "true" constraint $\widetilde{a}^T x \geq b$ at x^*. What happens is as follows:

• Over all vectors of coefficients $\widetilde{a}$ compatible with our 0.1%-uncertainty hypothesis, the minimum value of $\widetilde{a}^T x^* - b$, is < -104.9; in other words, the violation of the constraint can be as large as 450% of the right hand side!

• Treating the above worst-case violation as "too pessimistic" (why should the true values of all uncertain coefficients differ from the values indicated in (C) in the "most dangerous" way?), consider a less extreme measure of violation. Specifically, assume that the true values of the uncertain coefficients in (C) are obtained from the "nominal values" (those shown in (C)) by random perturbations $a_j \mapsto \widetilde{a}_j = (1+\xi_j)a_j$ with independent and, say, uniformly distributed on $[-0.001, 0.001]$ "relative perturbations" ξ_j. What will be a "typical" relative violation,

$$V = \max\left[\frac{b - \widetilde{a}^T x^*}{b}, 0\right] \times 100\%,$$

of the "true" (now random) constraint $\widetilde{a}^T x \geq b$ at x^*? The answer is nearly as bad as for the worst scenario:

Prob$\{V > 0\}$	Prob$\{V > 150\%\}$	Mean(V)
0.50	0.18	125%

Table 1. Relative violation of constraint 372 in `PILOT4`
(1,000-element sample of 0.1% perturbations of the uncertain data)

We see that *quite small (just 0.1%) perturbations of "obviously uncertain" data coefficients can make the "nominal" optimal solution x^* heavily infeasible and thus practically meaningless.*

A "case study" reported in [7] shows that the phenomenon we have just described is not an exception – in 13 of 90 *NETLIB* Linear Programming problems considered in this study, already 0.01%-perturbations of "ugly" coefficients result in violations of some constraints, as evaluated at the nominal optimal solutions by more than 50%. In 6 of these 13 problems the magnitude of constraint violations was over 100%, and in `PILOT4` — "the champion" — it was as large as 210,000%, that is, 7 orders of magnitude larger than the relative perturbations in the data.

> The techniques presented in this book as applied to the `NETLIB` problems allow one to eliminate the outlined phenomenon by passing out of the nominal optimal to *robust optimal* solutions. At the 0.1%-uncertainty level, the price of this "immunization against uncertainty" (the increase in the value of the objective when passing from the nominal to the robust solution), *for every one of the* `NETLIB` *problems*, is less than 1% (see [7] for details).

The outlined case study and many other examples lead to several observations:

A. *The data of real-world optimization problems more often than not are uncertain — not known exactly at the time the problem is being solved.* The reasons for data uncertainty include, among others:

measurement/estimation errors coming from the impossibility to measure/estimate exactly the data entries representing characteristics of physical systems/technological processes/environmental conditions, etc.

implementation errors coming from the impossibility to implement a solution exactly as it is computed. For example, whatever the entries "in reality" in the above nominal solution x^* to `PILOT4` — control inputs to physical systems, resources allocated for various purposes, etc. — they clearly cannot be implemented with the same high precision to which they are computed. The effect of the implementation errors, like $x_j^* \mapsto (1+\epsilon_j)x_j^*$, is as if there were no implementation errors, but the coefficients a_{ij} in the constraints of `PILOT4` were subject to perturbations $a_{ij} \mapsto (1+\epsilon_j)a_{ij}$.

B. *In real-world applications of Optimization one cannot ignore the possibility that even a small uncertainty in the data can make the nominal optimal solution to the problem completely meaningless from a practical viewpoint.*

C. *Consequently, in Optimization, there exists a real need of a methodology capable of detecting cases when data uncertainty can heavily affect the quality of the nominal solution, and in these cases to generate a robust solution, one that is immunized against the effect of data uncertainty.*

A methodology addressing the latter need is offered by *Robust Optimization*, which is the subject of this book.

B. Robust Optimization — The Paradigm

To explain the paradigm of Robust Optimization, we start by addressing the particular case of Linear Programming — the generic optimization problem that is perhaps the best known and the most frequently used in applications. Aside from its importance, this generic problem is especially well-suited for our current purposes, since the structure and the data of a Linear Programming program $\min\limits_x\{c^Tx : Ax \le b\}$ are clear. Given the form in which we wrote the program down, the structure is the sizes of the constraint matrix A, while the data is comprised of the numerical values of the entries in (c, A, b). In Robust Optimization, an *uncertain* LP problem is defined as a collection $\left\{\min_x\{c^Tx : Ax \le b\} : (c, A, B) \in \mathcal{U}\right\}$

of LP programs of a common structure with the data (c, A, b) varying in a given *uncertainty set* $\mathcal{U}$. The latter summarizes all information on the "true" data that is available to us when solving the problem. Conceptually, the most important question is what does it mean to solve an uncertain LP problem. The answer to this question, as offered by Robust Optimization in its most basic form, rests on three implicit assumptions on the underlying "decision-making environment":

A.1. All entries in the decision vector x represent "here and now" decisions: they should get specific numerical values as a result of solving the problem *before* the actual data "reveals itself."

A.2. The decision maker is fully responsible for consequences of the decisions to be made when, and only when, the actual data is within the prespecified uncertainty set $\mathcal{U}$.

A.3. The constraints of the uncertain LP in question are "hard" — the decision maker cannot tolerate violations of constraints when the data is in $\mathcal{U}$.

These assumptions straightforwardly lead to the definition of an "immunized against uncertainty" solution to an uncertain problem. Indeed, by A.1, such a solution should be a fixed vector that, by A.2 – A.3, should remain feasible for the constraints, whatever the realization of the data within $\mathcal{U}$; let us call such a solution *robust feasible*. Thus, in our decision-making environment, meaningful solutions to an uncertain problem are exactly its robust feasible solutions. It remains to decide how to interpret the value of the objective, (which can also be uncertain), at such a solution. As applied to the objective, our "worst-case-oriented" philosophy makes it natural to quantify the quality of a robust feasible solution x by the *guaranteed* value of the original objective, that is, by its largest value $\sup\left\{c^T x : (c, A, b) \in \mathcal{U}\right\}$. Thus, the best possible robust feasible solution is the one that solves the optimization problem

$$\min_x \left\{ \sup_{(c,A,b)\in\mathcal{U}} c^T x : Ax \le b \ \forall (c, A, b) \in \mathcal{U} \right\},$$

or, which is the same, the optimization problem

$$\min_{x,t} \left\{t : c^T x \le t,\ Ax \le b \ \forall (c, A, b) \in \mathcal{U}\right\}. \tag{RC}$$

The latter problem is called the *Robust Counterpart* (RC) of the original uncertain problem. The feasible/optimal solutions to the RC are called *robust feasible/robust optimal* solutions to the uncertain problem. The Robust Optimization methodology, in its simplest version, proposes to associate with an uncertain problem its Robust Counterpart and to use, as our "real life" decisions, the associated robust optimal solutions.

At this point, it is instructive to compare the RO paradigm with more traditional approaches to treating data uncertainty in Optimization, specifically, with *Stochastic Optimization* and *Sensitivity Analysis*.

Robust vs. Stochastic Optimization. In Stochastic Optimization (SO), the uncertain numerical data are assumed to be *random*. In the simplest case, these random data obey a known in advance probability distribution, while in more advanced settings, this distribution is only partially known. Here again an uncertain LP problem is associated with a deterministic counterpart, most notably with the *chance constrained* problem[1]

$$\min_{x,t}\left\{t : \mathrm{Prob}_{(c,A,b)\sim P}\left\{c^T x \leq t \ \&\ Ax \leq b\right\} \geq 1-\epsilon\right\}, \tag{ChC}$$

where $\epsilon \ll 1$ is a given tolerance and P is the distribution of the data (c, A, b). When this distribution is only partially known — all we know is that P belongs to a given family $\mathcal{P}$ of probability distributions on the space of the data — the above setting is replaced with the *ambiguous chance constrained* setting,

$$\min_{x,t}\left\{t : \mathrm{Prob}_{(c,A,b)\sim P}\left\{c^T x \leq t \ \&\ Ax \leq b\right\} \geq 1-\epsilon \,\forall P \in \mathcal{P}\right\}. \tag{Amb}$$

The SO approach seems to be less conservative than the worst-case-oriented RO approach. However, this is so *if* indeed the uncertain data are of a stochastic nature, *if* we are smart enough to point out the associated probability distribution (or at least a "narrow" family of distributions to which the true one belongs), and *if* indeed we are ready to accept probabilistic guarantees as given by chance constraints. The three if's above are indeed satisfied in some applications, such as Signal Processing, or analysis and synthesis of service systems[2]. At the same time, in numerous applications the three aforementioned if's are too restrictive. Think, e.g., of measurement/estimation errors for *individual* problems, like `PILOT4`. Even assuming that preparation of data entries for `PILOT4` indeed involved something random, we perhaps could think about the distribution of the nominal data given the true ones, but not about what we actually need — the distribution of the true data given the nominal ones. The latter most probably just does not make sense — `PILOT4` represents a particular decision-making problem with particular deterministic (albeit not known to us exactly) data, and all we can say about this true data given the nominal ones, is that the former data lies in given confidence intervals around the nominal data (and even this can be said under the assumption that when

[1] The concept of chance constraints goes back to A. Charnes, W.W. Copper, and G.H. Symonds [40], 1958. An alternative to chance constrained setting is where we want to optimize the expected value of the objective (the latter can incorporate penalty terms for violation of uncertain constraints) under the certain part of the original constraints. This approach, however, is aimed at "soft" constraints, while we are primarily interested in the case there the constraints are hard.

[2] Indeed, in these subject areas the random factors (like observation noises in Signal Processing, or interarrival/service times in service systems) are of random nature with more or less easy-to-identify distributions, especially when we have reasons to believe that different components of random data (like different entries in the observation noises, or individual inter-arrival and service times) are independent of each other. In such situations identifying the distribution of the data reduces to identifying a bunch of low-dimensional distributions, which is relatively easy. Furthermore, the systems in question are aimed at servicing many customers over long periods of time, so that here the probabilistic guarantees do make sense. For example, day by day many hundreds/thousands of users are sending/receiving e-mails or contacting a calling center, and a probabilistic description of the service level (the probability for an e-mail to be lost, or for the time to get an operator response to become unacceptably long) makes good sense — it merely says that in the long run, a certain fraction of users/custmers will be dissatisfied.

measuring the true data to get the nominal, no "rare event" took place). Further, even when the true data indeed are of a stochastic nature, it is usually difficult to properly identify the underlying distributions. Unless there are good reasons to a priori specify these distributions up to a small number of parameters that further can be estimated sufficiently well from observations[3], accurate identification of a "general type" multi-dimensional probability distribution usually requires an astronomical, completely unrealistic number of observations. As a result, Stochastic Optimization more often than not is forced to operate with oversimplified *guesses* for the actual distributions (like the log-normal factor model for stock returns), and usually it is very difficult to evaluate the influence of this new uncertainty — in the probability distribution — on the quality of the SO-based decisions.

The third of the above if's, our willingness to accept probabilistic guarantees, also can be controversial. Imagine, for the sake of argument, that we have at our disposal a perfect stochastic model of the stock market — as solid as the transparent model of a lottery played every week in many countries. Does the relevance of the stochastic model of the stock market make the associated probabilistic guarantees of the performance of a pension fund really meaningful for an individual customer, as meaningful as a similar guarantee in the lottery case? We believe that many customers will answer this question negatively, and rightfully so. People playing a lottery on a regular basis during their life span, participate in several hundreds of lotteries, and thus can refer to the Law of Large Numbers as a kind of indication that probabilistic guarantees indeed are meaningful for them. In contrast to this, every individual plays the "pension fund lottery" just once, which makes the interpretation of probabilistic guarantees much more problematic. Of course, the three if's above become less restrictive when passing from the chance constrained problem (ChC), where the distribution of the uncertain data is known exactly, to the ambiguously chance constrained problem (Amb), and become the less restrictive the wider families of distributions $\mathcal{P}$ we are ready to consider. Note, however, that passing from (ChC) to (Amb) is, conceptually, a step towards the Robust Counterpart — the latter is nothing but the ambiguously chance constrained problem associated with the family $\mathcal{P}$ of *all* distributions supported on a given set $\mathcal{U}$.

In fact the above three if's should be augmented by a fourth, even more restrictive "if" — chance constrained settings (ChC) and (Amb) can be treated as actual sources of "immunized against uncertainty" decisions only *if* these problems are computationally tractable; when that is not the case, these settings become more wishful thinking than actual decision-making tools. As a matter of fact, the computational tractability of chance constrained problems is a pretty rare commodity — aside of a number of very particular cases, it is difficult to verify (especially when ϵ is really small) whether a given candidate solution is feasible for a chance constrained problem. In addition, chance constraints more often than not result in *nonconvex* feasible sets, which make the optimization required in (ChC) and (Amb)

[3]For example, one can refer to the Central Limit Theorem in order to justify the standard — the Gaussian — model of noise in communications.

highly problematic. In sharp contrast to this, the Robust Counterparts of uncertain *Linear Programming* problems are *computationally tractable*, provided the underlying uncertainty sets $\mathcal{U}$ satisfy mild convexity and computability assumptions (e.g., are given by explicit systems of efficiently computable convex inequalities).

It should be added that the "conservatism" of RO as compared to SO is in certain respects an advantage rather than a disadvantage. When designing a construction, like a railroad bridge, by applying quantitative techniques, engineers usually increase the safety-related design parameters, like thicknesses of bars, by a reasonable margin, such as 30 to 50%, in order to account for modeling inaccuracies, rare but consequential environmental conditions, etc. With the Robust Optimization approach, this desire "to stay on the safe side" can be easily achieved by enlarging the uncertainty set. This is not the case in a chance constrained problem (ChC), where the total "budget of uncertainty" is fixed — the total probability mass of all realizations of the uncertain data must be one, so that when increasing the probabilities of some "scenarios" to make them more "visible," one is forced to reduce probabilities of other scenarios, and there are situations where this phenomenon is difficult to handle. Here again, in order to stay "on the safe side" one needs to pass from chance constrained problems to their ambiguously chance constrained modifications, that is, to move towards Robust Counterparts.

In our opinion, Stochastic and Robust Optimization are *complementary* approaches for handling data uncertainty in Optimization, each having its own advantages and drawbacks. For example, information on the stochastic nature of data uncertainty, if any, can be utilized in the RO framework, as a kind of a guideline for building uncertainty sets $\mathcal{U}$. It turns out that the latter can be built in such a way that by immunizing a candidate solution against *all* realizations of the data from $\mathcal{U}$, we automatically immunize it against *nearly all* (namely, up to realizations of total probability mass $\leq \epsilon$) random perturbations, thus making the solution feasible for the chance constrained problem. A naive way to achieve this goal would be to choose $\mathcal{U}$ as a computationally tractable convex set that "$(1-\epsilon)$-supports" all distributions from $\mathcal{P}$ (that is, $P(\mathcal{U}) \geq 1-\epsilon$ for all $P \in \mathcal{P}$). In this book, however, we show that under mild assumptions there exist less evident and *incomparably less conservative* ways to come up with uncertainty sets achieving the above goal.

Robust Optimization and Sensitivity Analysis. Along with Stochastic Optimization, another traditional body of knowledge dealing, in a sense, with data uncertainty in optimization is *Sensitivity Analysis*. Here the issues of primary importance are the continuity properties of the usual (the nominal) optimal solution as a function of the underlying nominal data. It is immediately seen that both Robust and Stochastic Optimization are aimed at answering the same question (albeit in different settings), the question of building an uncertainty-immunized solution to an optimization problem with uncertain data; Sensitivity Analysis is aimed at a completely different question.

Robust Optimization History. Robust optimization has many roots and precursors in the applied sciences. Some of these connections are explicit, while others are a way, looking backwards in time, to interpret an approach that was developed under different ideas. We mention three areas where robustness has played, and continues to play, an important role.

Robust Control. The field of *Robust Control* has evolved, mainly during the 90s, in the interest of control systems designers for some level of guarantee in terms of stability of the controlled system. The quest for robustness can be historically traced back to the concept of a stability margin developed in the early 30s by Bode and others, in the context of feedback amplifiers. Questions such as the "stability margin," which is the amount of feedback gain required to de-stabilize a controlled system, led naturally to a "worst-case" point of view, in which "bad" parameter values are too dangerous to be allowed, even with low probability. In the late 80s, the then-classical approach to control of large-scale feedback systems, which was based on stochastic optimization ideas, came under criticism as it could not be guaranteed to offer any kind of stability margin. The approach of $\mathbf{H}_\infty$ control was then developed as a multivariate generalization of the stability margin in the early 90s. Later, the approach was extended under the name μ-control, to handle more general, parametric perturbations (the $\mathbf{H}_\infty$ norm measures robustness with respect to a very special kind of perturbation). The corresponding robust control design problem turns out to be difficult, but relaxations based on conic (precisely, semidefinite) optimization were introduced under the name of Linear Matrix Inequalities.

Robust Statistics. In Statistics, robustness usually refers to insensitivity to outliers. Huber (see, e.g., [65]) has proposed a way to handle outliers by a modification of loss functions. The precise connection with Robust Optimization is yet to be made.

Machine Learning. More recently, the field of Machine Learning has witnessed great interest in Support Vector Machines, which are classification algorithms that can be interpreted as maximizing robustness to a special kind of uncertainty. We return to this topic in chapter 12.

Robust Linear and Convex Optimization. Aside of the outlined precursors, the paradigm of Robust Optimization per se, in the form considered here, goes back to A.L. Soyster [109], who was the first to consider, as early as in 1973, what now is called Robust Linear Programming. To the best of our knowledge, in two subsequent decades there were only two publications on the subject [52, 106]. The activity in the area was revived circa 1997, independently and essentially simultaneously, in the frameworks of both Integer Programming (Kouvelis and Yu [70]) and

Convex Programming (Ben-Tal and Nemirovski [3, 4], El Ghaoui et al. [49, 50]). Since 2000, the RO area is witnessing a burst of research activity in both theory and applications, with numerous researchers involved worldwide. The magnitude and diversity of the related contributions make it beyond our abilities to discuss them here. The reader can get some impression of this activity from [9, 16, 110, 89] and references therein.

C. The Scope of Robust Optimization and Our Focus in this Book

By itself, the RO methodology can be applied to every generic optimization problem where one can separate numerical data (that can be partly uncertain and are only known to belong to a given uncertainty set) from problem's structure (that is known in advance and is common for all instances of the uncertain problem). In particular, the methodology is fully applicable to

- *conic problems* — convex problems of the form

$$\min_x \left\{c^T x : b - Ax \in \mathbf{K}\right\}, \tag{C}$$

  where $\mathbf{K}$ is a given "well-structured" convex cone, representing, along with the sizes of A, a problem's structure, while the numerical entries (c, A, b) form problems's data. Conic problems look very similar to LP programs that are recovered when $\mathbf{K}$ is specified as the nonnegative orthant $\mathbb{R}^m_+$. Two other common choices of the cone $\mathbf{K}$ are:

  — a direct product of *Lorentz cones* of different dimensions. The k-dimensional Lorentz cone (also called the Second Order, or the Ice-Cream cone) is defined as

$$\mathbf{L}^k = \{x \in \mathbb{R}^k : x_k \geq \sqrt{x_1^2 + ... + x_{k-1}^2}.$$

  Problems (C) with direct products of Lorentz cones in the role of $\mathbf{K}$ are called *Conic Quadratic*, or *Second Order Conic* Programming problems;

  — a direct product of *semidefinite cones* of different sizes. The semidefinite cone of size k, denoted by $\mathbf{S}^k_+$, is the set of all symmetric positive semidefinite $k \times k$ matrices; it "lives" in the linear space $\mathbf{S}^k$ of all symmetric $k \times k$ matrices. Problems (C) with direct products of semidefinite cones in the role of $\mathbf{K}$ are called *Semidefinite programs*.

  Conic Quadratic and especially Semidefinite Programming problems possess extremely rich "expressive abilities"; in fact, Semidefinite Programming "captures" nearly all convex problems arising in applications, see, e.g., [8, 32, 33].

— *Integer and Mixed Integer Linear Programming* – Linear Programming problems where all or part of the variables are further restricted to be integers.

The broad spectrum of research questions related to Robust Optimization can be split into three main categories.

i) *Extensions of the RO paradigm.* It turns out that the implicit assumptions A.1, A.2, A.3 that led us to the central notion of Robust Counterpart of an uncertain optimization problem, while meaningful in numerous applications, in some other applications do not fully reflect the possibilities to handle the uncertain data. At present, two extensions addressing this added flexibility exist:

• *Globalized Robust Counterpart.* This extension of the notion of RC corresponds to the case when we revise Assumption A.2. Specifically, now we require a candidate solution x to satisfy the constraints for all instances of the uncertain data in $\mathcal{U}$ and, in addition, seek for *controlled* deterioration of the constraints evaluated at x when the uncertain data runs out of $\mathcal{U}$. The corresponding analogy to the Robust Counterpart of an uncertain (conic) problem, called *Globalized Robust Counterpart* (GRC), is the optimization program

$$\min_{x,t}\left\{t:\begin{array}{l}c^Tx-t\leq\alpha_{\rm obj}{\rm dist}((c,A,b),\mathcal{U})\\{\rm dist}(b-Ax,\mathbf{K})\leq\alpha_{\rm cons}{\rm dist}((c,A,b),\mathcal{U})\end{array}\right\}\ \forall(c,A,b)\Bigg\},\qquad\text{(GRC)}$$

where the distances come from given norms on the corresponding spaces and $\alpha_{\rm obj}$, $\alpha_{\rm cons}$ are given nonnegative *global sensitivities.*

• *Adjustable Robust Counterpart.* This extension of the notion of RC corresponds to the case when some of the decision variables x_j represent "wait and see" decisions to be made when the true data partly reveals itself, or are *analysis variables* that do not represent decisions (e.g., slack variables introduced to convert the original problem into a prescribed form, say, an LP one). It is natural to allow these *adjustable* variables to adjust themselves to the true data. Specifically, we can assume that every decision variable x_j is allowed to depend on a given "portion" $P_j(c,A,b)$ of the true data (c,A,B) of a (conic) problem:

$$x_j=X_j(P_j(c,A,b)),$$

where $X_j(\cdot)$ can be arbitrary functions. We then require from the resulting *decision rules* to satisfy the constraints of the uncertain conic problem for all realizations of the data from $\mathcal{U}$. The corresponding *Adjustable Robust Counterpart* (ARC) of an uncertain conic problem is the optimization program

$$\min_{X_1(\cdot),\ldots,X_n(\cdot),t}\left\{t:b-A\begin{bmatrix}X_1(P_1(c,A,b))\\\vdots\\X_n(P_n(c,A,b))\end{bmatrix}\in\mathbf{K}\ \forall(c,A,b)\in\mathcal{U}\right\}.\qquad\text{(ARC)}$$

It should be stressed that the optimization in (ARC) is carried out not over finite-dimensional vectors, as is the case in RC and GRC, but over infinite-dimensional *decision rules* — arbitrary functions $X_j(\cdot)$ on the corresponding finite-dimensional vector spaces. In order to cope, to some extent, with a severe computational intractability of ARCs, one can restrict the structure of decision rules, most notably, to make them affine in their arguments:

$$X_j(p_j)=q_j+r_j^Tp_j.$$

When restricted to affine decision rules, the ARC becomes an optimization problem in finitely many real variables q_j, r_j, $1 \le j \le n$. This problem is called the *Affinely Adjustable Robust Counterpart* (AARC) of the original uncertain conic problem corresponding to the *information base* $P_1(\cdot), ..., P_n(\cdot)$.

ii) *Investigating tractability issues of Robust Counterparts.* Already the plain Robust Counterpart,

$$\min_{x,t} \left\{ t : c^T x \le t, b - Ax \in \mathbf{K} \ \forall (c, A, b) \in \mathcal{U} \right\}, \tag{RC}$$

of an uncertain conic problem,

$$\left\{ \min_{x} \left\{ c^T x : b - Ax \in \mathbf{K} \right\} : (c, A, b) \in \mathcal{U} \right\},$$

has a more complicated structure than an instance of the uncertain problem itself: (RC) is what is called a *semi-infinite* conic problem, one with *infinitely many* conic constraints $\begin{bmatrix} t - c^T x \\ b - Ax \end{bmatrix} \in \mathbf{K}_+ = \mathbb{R}_+ \times \mathbf{K}$ parameterized by the uncertain data (c, A, b) running through the uncertainty set. While (RC) is still convex, the semi-infinite nature makes it more difficult computationally than the instances of the associated uncertain problem. It may well happen that (RC) is computationally intractable, even when the uncertainty set $\mathcal{U}$ is a nice convex set (say, a ball, or a polytope) and the cone $\mathbf{K}$ is as simple as in the case of Conic Quadratic and Semidefinite programs. At the same time, in order for RO to be a working tool rather than wishful thinking, we need the RC to be computationally tractable; after all, what is the point in reducing something to an optimization problem that we do not know how to process computationally? This motivates what is in our opinion *the* the main theoretical challenge in Robust Optimization: *identifying the cases where the RC (GRC, AARC, ARC) of an uncertain conic problem admits a computationally tractable equivalent reformulation, or at least a computationally tractable safe approximation.* Here safety means that every feasible solution to the approximation is feasible for the "true" Robust Counterpart.

At the present level of our knowledge, the "big picture" here is as follows.
• When the cone $\mathbf{K}$ is "as simple as possible," i.e., is a nonnegative orthant (the case of uncertain Linear Programming), the Robust Counterpart (and under mild additional structural conditions, the GRC and the AARC as well) is computationally tractable, provided that the underlying (convex) uncertainty set $\mathcal{U}$ is so. The latter means that $\mathcal{U}$ is a convex set given by an explicit system of efficiently computable convex constraints (say, a polytope given by an explicit list of linear inequalities).
• When the (convex) uncertainty set $\mathcal{U}$ is "as simple as possible," i.e., a polytope given as a convex hull of a finite set of reasonable cardinality (scenario uncertainty), the RC is computationally tractable whenever $\mathbf{K}$ is a computationally tractable convex cone, as is the case in Linear, Conic Quadratic, and Semidefinite Programming.

• In between the above two extremes, for example, in the case of uncertain Conic Quadratic and Semidefinite problems with polytopes in the role of uncertainty sets, the RCs are, in general, computationally intractable. There are however particular cases, important for applications, where the RC is tractable, and even more cases where it admits safe tractable approximations that are *tight*, in a certain precise sense.

iii) *Applications.* This avenue of RO research is aimed at building and processing Robust Counterparts of specific optimization problems arising in various applications.

The position of our book with respect to these three major research areas in Robust Optimization is as follows:

> *Our primary emphasis is on presenting in full detail the Robust Optimization paradigm* (including its recent extensions mentioned in item 1, as well as links with Chance Constrained Stochastic Optimization) *and tractability issues, primarily for Uncertain Linear, Conic Quadratic, and Semidefinite Programming.*

D. Prerequisites and Contents

Prerequisites for reading this book are quite modest — essentially, all that is expected of a reader is knowledge of basic Analysis, Linear Algebra, and Probability, plus general mathematical culture. Preliminary "subject-specific" knowledge, (which in our case means knowledge of the Convex Optimization basics, primarily of Conic Programming and Conic Duality, on one hand, and of tractability issues in Convex Programming, on the other), while being highly welcomed, is not absolutely necessary. All required basics can be found in the Appendix augmenting the main body of the book.

The contents. The main body of the book is split into four parts:

• Part I is the basic theory of the "here and now" (i.e., the non-adjustable) Robust Linear Programming, which starts with detailed discussion of the concepts of an uncertain Linear Programming problem and its Robust/Generalized Robust Counterparts. Along with other results, we demonstrate that the RC/GRC of an uncertain LP problem is computationally tractable, provided that the uncertainty set is so. As it was already mentioned, such a general tractability result is a specific feature of uncertain LP. Another major theme of Part I is that of computationally tractable safe approximations of chance constrained uncertain LP problems with randomly perturbed data.

Part I, perhaps with chapter 4 skipped, can be used as a stand-alone graduate-level textbook on Robust Linear Programming, or as a base of a semester-long graduate course on Robust Optimization.

• Part II can be treated as a "conic version" of Part I, where the main concepts of non-adjustable Robust Optimization are extended to uncertain Convex Programming problems in the conic form, with emphasis on uncertain Conic Quadratic and Semidefinite Programming problems. As it was already mentioned, aside of the (in fact, trivial) case of scenario uncertainty, Robust/Generalized Robust Counterparts of uncertain CQP/SDP problems are, in general, computationally intractable. This is why the emphasis is on identifying, and illustrating the importance of generic situations where the RCs/GRCs of uncertain Conic Quadratic/Semidefinite problems admit tractable reformulation, or a tight safe tractable approximation. Another theme considered in Part II is that of safe tractable approximation of chance constrained uncertain Conic Quadratic and Semidefinite problems with randomly perturbed data. As compared to its "LP predecessor" from Part I, this theme now has an unexpected twist: it turns out that safe tractable approximations of the chance constrained Conic Quadratic/Semidefinite inequalities are easier to build and to process than the tight safe tractable approximations to the RCs of these conic inequalities. This is completely opposite of what happens in the case of uncertain LP problems, where it is easy to process *exactly* the RCs, but not the chance constrained versions of uncertain linear inequality constraints.

We conclude Part II investigating Robust Counterparts of specific "well structured" uncertain convex constraints arising in Machine Learning and Linear Regression models. Since the most interesting uncertain constraints arising in this context are neither Conic Quadratic nor Semidefinite, the tractability-related questions associated with the RCs of these constraints need a dedicated treatment, and this is our major goal in the corresponding chapter.

• Part III is devoted to *Robust Multi-Stage Decision Making*, specifically, to Robust Dynamic Programming, and to Adjustable (with emphasis on *Affinely* Adjustable) Robust Counterparts of uncertain conic problems, primarily uncertain multi-stage LPs. As always, our emphasis is on the tractability issues. We demonstrate, in particular, that the AARC methodology allows for efficient handling of the finite-horizon synthesis of linear controllers for uncertainty-affected Linear Dynamical systems with certain (and known in advance) dynamics. The design specifications in this synthesis can be given by general-type systems of linear constraints on states and controls, to be satisfied in a robust with respect to the initial state and the external inputs fashion.

• A short, single-chapter Part IV presents three realistic examples, worked out in full detail, of application of the RO methodology. While not pretending to give an impression of a wide and diverse range of existing applications of RO, these examples, we believe, add a "bit of reality" to our primarily theoretical treatment of the subject.

Reading modes. We believe that acquaintance with Part I is a natural prerequisite for reading Parts II and III; however, the latter two parts can be read independently of each other. In addition, those not interested in the theme of chance constraints, may skip the related chapters 2, 4 and 10; those interested in this theme may in the first reading skip chapter 4.

Acknowledgments. In our decade-long research on Robust Optimization, summarized in this book, we collaborated with many excellent colleagues and excellent students. This collaboration contributed significantly to the contents of this book. We are greatly indebted to those with whom we had the honor and privilege to collaborate in various ways in our RO-related research: A. Beck, D. Bertsimas, S. Boyd, O. Boni, D. Brown, G. Calafiore, M. Campi, F. Dan Barb, Y. Eldar, B. Golany, A. Goryashko, E. Guslitzer, R. Hildenbrand, G. Iyengar, A. Juditsky, M. Kočvara, H. Lebret, T. Margalit, Yu. Nesterov, A. Nilim, F. Oustry, C. Roos, A. Shapiro, S. Shtern, M. Sim, T. Terlaky, U. Topcu, L. Tuncel, J.-Ph. Vial, J. Zowe. Special gratitude goes to D. den Hertog and E. Stinstra for permitting us to use the results of their research on robust models in TV tube manufacturing (Section 15.1).

We gratefully acknowledge the relevant financial support from several agencies: Israeli Ministry of Science (grant 0200–1–98), Israeli Science Foundation (grants 306/94–3, 683/99–10.0), Germany-Israel Foundation for R&D (grant I0455–214.06/95), Minerva Foundation, Germany, US National Science Foundation (grants DMS-0510324, DMI-0619977, MSPA-MCS-0625371, SES-0835531), and US-Israel Binational Science Foundation (grant 2002038).

Aharon Ben-Tal
Laurent El Ghaoui
Arkadi Nemirovski
November 2008

Part I

Robust Linear Optimization

This page intentionally left blank

Chapter One

Uncertain Linear Optimization Problems and their Robust Counterparts

In this chapter, we introduce the concept of the uncertain Linear Optimization problem and its Robust Counterpart, and study the computational issues associated with the emerging optimization problems.

1.1 DATA UNCERTAINTY IN LINEAR OPTIMIZATION

Recall that the Linear Optimization (LO) *problem* is of the form

$$\min_x \left\{ c^T x + d : Ax \le b \right\}, \tag{1.1.1}$$

where $x \in \mathbb{R}^n$ is the vector of *decision variables*, $c \in \mathbb{R}^n$ and $d \in \mathbb{R}$ form the *objective*, A is an $m \times n$ *constraint matrix*, and $b \in \mathbb{R}^m$ is the *right hand side vector*.

> Clearly, the constant term d in the objective, while affecting the optimal value, does not affect the optimal solution, this is why it is traditionally skipped. As we shall see, when treating the LO problems with *uncertain data* there are good reasons not to neglect this constant term.

The *structure* of problem (1.1.1) is given by the number m of constraints and the number n of variables, while the *data* of the problem are the collection (c, d, A, b), which we will arrange into an $(m+1) \times (n+1)$ *data matrix*

$$D = \left[\begin{array}{c|c} c^T & d \\ \hline A & b \end{array} \right].$$

> Usually not all constraints of an LO program, as it arises in applications, are of the form $a^T x \le \text{const}$; there can be linear "$\ge$" inequalities and linear equalities as well. Clearly, the constraints of the latter two types can be represented equivalently by linear "$\le$" inequalities, and we will assume henceforth that these are the only constraints in the problem.

Typically, the data of real world LOs (Linear Optimization problems) is not known exactly. The most common reasons for data uncertainty are as follows:

- Some of data entries (future demands, returns, etc.) do not exist when the problem is solved and hence are replaced with their forecasts. These data entries are thus subject to *prediction errors*;

- Some of the data (parameters of technological devices/processes, contents associated with raw materials, etc.) cannot be measured exactly – in reality their values drift around the measured "nominal" values; these data are subject to *measurement errors*;
- Some of the decision variables (intensities with which we intend to use various technological processes, parameters of physical devices we are designing, etc.) cannot be implemented exactly as computed. The resulting *implementation errors* are equivalent to appropriate artificial data uncertainties.

  Indeed, the contribution of a particular decision variable x_j to the left hand side of constraint i is the product $a_{ij}x_j$. Hence the consequences of an additive implementation error $x_j \mapsto x_j + \epsilon$ are as if there were no implementation error at all, but the left hand side of the constraint got an extra additive term $a_{ij}\epsilon$, which, in turn, is equivalent to the perturbation $b_i \mapsto b_j - a_{ij}\epsilon$ in the right hand side of the constraint. The consequences of a more typical *multiplicative* implementation error $x_j \mapsto (1+\epsilon)x_j$ are as if there were no implementation error, but each of the data coefficients a_{ij} was subject to perturbation $a_{ij} \mapsto (1+\epsilon)a_{ij}$. Similarly, the influence of additive and multiplicative implementation error in x_j on the value of the objective can be mimicked by appropriate perturbations in d or c_j.

In the traditional LO methodology, a small data uncertainty (say, 1% or less) is just ignored; the problem is solved *as if* the given ("nominal") data were exact, and the resulting *nominal* optimal solution is what is recommended for use, in hope that small data uncertainties will not affect significantly the feasibility and optimality properties of this solution, or that small adjustments of the nominal solution will be sufficient to make it feasible. We are about to demonstrate that these hopes are not necessarily justified, and sometimes even small data uncertainty deserves significant attention.

1.1.1 Introductory Example

Consider the following very simple linear optimization problem:

Example 1.1.1. A company produces two kinds of drugs, DrugI and DrugII, containing a specific active agent A, which is extracted from raw materials purchased on the market. There are two kinds of raw materials, RawI and RawII, which can be used as sources of the active agent. The related production, cost, and resource data are given in table 1.1. The goal is to find the production plan that maximizes the profit of the company.

Parameter	DrugI	DrugII
Selling price, $ per 1000 packs	6,200	6,900
Content of agent A, g per 1000 packs	0.500	0.600
Manpower required, hours per 1000 packs	90.0	100.0
Equipment required, hours per 1000 packs	40.0	50.0
Operational costs, $ per 1000 packs	700	800

(a) Drug production data

Raw material	Purchasing price, $ per kg	Content of agent A, g per kg
RawI	100.00	0.01
RawII	199.90	0.02

(b) Contents of raw materials

Budget, $	Manpower, hours	Equipment, hours	Capacity of raw materials storage, kg
100,000	2,000	800	1,000

(c) Resources

Table 1.1 Data for Example 1.1.1.

The problem can be immediately posed as the following linear programming program:

(Drug):

$$\begin{aligned} \text{Opt} \;=\; \min\Big\{ & \overbrace{[100\cdot \mathit{RawI} + 199.90\cdot \mathit{RawII} + 700\cdot \mathit{DrugI} + 800\cdot \mathit{DrugII}]}^{\text{purchasing and operational costs}} \\ & - \underbrace{[6200\cdot \mathit{DrugI} + 6900\cdot \mathit{DrugII}]}_{\text{income from selling the drugs}} \Big\} \qquad \text{[minus total profit]} \end{aligned}$$

subject to

$$\begin{array}{lr} 0.01\cdot \mathit{RawI} + 0.02\cdot \mathit{RawII} - 0.500\cdot \mathit{DrugI} - 0.600\cdot \mathit{DrugII} \ge 0 & \text{[balance of active agent]} \\ \mathit{RawI} + \mathit{RawII} \le 1000 & \text{[storage constraint]} \\ 90.0\cdot \mathit{DrugI} + 100.0\cdot \mathit{DrugII} \le 2000 & \text{[manpower constraint]} \\ 40.0\cdot \mathit{DrugI} + 50.0\cdot \mathit{DrugII} \le 800 & \text{[equipment constraint]} \\ 100.0\cdot \mathit{RawI} + 199.90\cdot \mathit{RawII} + 700\cdot \mathit{DrugI} + 800\cdot \mathit{DrugII} \le 100000 & \text{[budget constraint]} \\ \mathit{RawI}, \mathit{RawII}, \mathit{DrugI}, \mathit{DrugII} \ge 0 & \end{array}$$

The problem has four variables — the amounts *RawI*, *RawII* (in kg) of raw materials to be purchased and the amounts *DrugI*, *DrugII* (in 1000 of packs) of drugs to be produced.

The optimal solution of our LO problem is

$$\text{Opt} = -8819.658; \mathit{RawI} = 0, \mathit{RawII} = 438.789, \mathit{DrugI} = 17.552, \mathit{DrugII} = 0.$$

Note that both the budget and the balance constraints are active (that is, the production process utilizes the entire 100,000 budget and the full amount of active agent contained in the raw materials). The solution promises the company a modest, but quite respectable profit of 8.8%.

1.1.2 Data Uncertainty and its Consequences

Clearly, even in our simple problem some of the data cannot be "absolutely reliable"; e.g., one can hardly believe that the contents of the active agent in the raw materials are exactly 0.01 g/kg for RawI and 0.02 g/kg for RawII. In reality, these contents vary around the indicated values. A natural assumption here is that the actual contents of active agent aI in RawI and aII in RawII are realizations of random variables somehow distributed around the "nominal contents" $anI = 0.01$ and $anII = 0.02$. To be more specific, assume that aI drifts in a 0.5% margin of anI, thus taking values in the segment $[0.00995, 0.01005]$. Similarly, assume that aII drifts in a 2% margin of $anII$, thus taking values in the segment $[0.0196, 0.0204]$. Moreover, assume that aI, aII take the two extreme values in the respective segments with probabilities 0.5 each. How do these perturbations of the contents of the active agent affect the production process? The optimal solution prescribes to purchase 438.8 kg of RawII and to produce 17.552K packs of DrugI (K stands for "thousand"). With the above random fluctuations in the content of the active agent in RawII, this production plan will be infeasible with probability 0.5, i.e., the actual content of the active agent in raw materials will be less than the one required to produce the planned amount of DrugI. This difficulty can be resolved in the simplest way: when the actual content of the active agent in raw materials is insufficient, the output of the drug is reduced accordingly. With this policy, the actual production of DrugI becomes a random variable that takes with equal probabilities the nominal value of 17.552K packs and the (2% less) value of 17.201K packs. These 2% fluctuations in the production affect the profit as well; it becomes a random variable taking, with probabilities 0.5, the nominal value 8,820 and the 21% (!) less value 6,929. The expected profit is 7,843, which is 11% less than the nominal profit 8,820 promised by the optimal solution of the nominal problem.

We see that in our simple example a pretty small (and unavoidable in reality) perturbation of the data may make the nominal optimal solution infeasible. Moreover, a straightforward adjustment of the nominally optimal solution to the actual data may heavily affect the quality of the solution.

Similar phenomenon can be met in many practical linear programs where at least part of the data are not known exactly and can vary around their nominal values. The consequences of data uncertainty can be much more severe than in our toy example. The analysis of linear optimization problems from the NETLIB collection[1] reported in [7] reveals that for 13 of 94 NETLIB problems, random 0.01% perturbations of the uncertain data can make the nominal optimal solution severely infeasible: with a non-negligible probability, it violates some of the constraints by

[1] A collection of LP programs, including those of real world origin, used as a standard benchmark for testing LP solvers.

50% and more. It should be added that in the general case (in contrast to our toy example) there is no evident way to adjust the optimal solution to the actual values of the data by a small modification, and there are cases when such an adjustment is in fact impossible; in order to become feasible for the perturbed data, the nominal optimal solution should be "completely reshaped."

The conclusion is as follows:

In applications of LO, there exists a real need of a technique capable of detecting cases when data uncertainty can heavily affect the quality of the nominal solution, and in these cases to generate a "reliable" solution, one that is immunized against uncertainty.

We are about to introduce the *Robust Counterpart* approach to uncertain LO problems aimed at coping with data uncertainty.

1.2 UNCERTAIN LINEAR PROBLEMS AND THEIR ROBUST COUNTERPARTS

Definition 1.2.1. An uncertain Linear Optimization problem is a collection

$$\left\{\min_x \left\{c^T x + d : Ax \le b\right\}\right\}_{(c,d,A,b)\in\mathcal{U}} \qquad (\mathrm{LO}_{\mathcal{U}})$$

of LO problems (instances) $\min_x \left\{c^T x + d : Ax \le b\right\}$ of common structure (i.e., with common numbers m of constraints and n of variables) with the data varying in a given uncertainty set $\mathcal{U} \subset \mathbb{R}^{(m+1)\times(n+1)}$.

We always assume that the uncertainty set is parameterized, in an affine fashion, by *perturbation vector* ζ varying in a given *perturbation set* $\mathcal{Z}$:

$$\mathcal{U} = \left\{ \left[\begin{array}{c|c} c^T & d \\ \hline A & b \end{array}\right] = \underbrace{\left[\begin{array}{c|c} c_0^T & d_0 \\ \hline A_0 & b_0 \end{array}\right]}_{\substack{\text{nominal} \\ \text{data } D_0}} + \sum_{\ell=1}^{L} \zeta_\ell \underbrace{\left[\begin{array}{c|c} c_\ell^T & d_\ell \\ \hline A_\ell & b_\ell \end{array}\right]}_{\substack{\text{basic} \\ \text{shifts } D_\ell}} : \zeta \in \mathcal{Z} \subset \mathbb{R}^L \right\}. \qquad (1.2.1)$$

For example, the story told in section 1.1.2 makes (Drug) an uncertain LO problem as follows:

- *Decision vector:* $x = [RawI; RawII; DrugI; DrugII]$;
- *Nominal data:* $D_0 = \left[\begin{array}{cccc|c} 100 & 199.9 & -5500 & -6100 & 0 \\ \hline -0.01 & -0.02 & 0.500 & 0.600 & 0 \\ 1 & 1 & 0 & 0 & 1000 \\ 0 & 0 & 90.0 & 100.0 & 2000 \\ 0 & 0 & 40.0 & 50.0 & 800 \\ 100.0 & 199.9 & 700 & 800 & 100000 \\ -1 & 0 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 & 0 \\ 0 & 0 & -1 & 0 & 0 \\ 0 & 0 & 0 & -1 & 0 \end{array}\right]$

- *Two basic shifts:*

$$D_1 = 5.0\text{e-}5 \cdot \left[\begin{array}{cccc|c} 0 & 0 & 0 & 0 & 0 \\ \hline 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{array}\right], \quad D_2 = 4.0\text{e-}4 \cdot \left[\begin{array}{cccc|c} 0 & 0 & 0 & 0 & 0 \\ \hline 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{array}\right]$$

- *Perturbation set:*

$$\mathcal{Z} = \left\{\zeta \in \mathbb{R}^2 : -1 \leq \zeta_1, \zeta_2 \leq 1\right\}.$$

This description says, in particular, that the only uncertain data in (Drug) are the coefficients *anI*, *anII* of the variables *RawI*, *RawII* in the balance inequality, (which is the first constraint in (Drug)), and that these coefficients vary in the respective segments $[0.01 \cdot (1 - 0.005), 0.01 \cdot (1 + 0.005)]$, $[0.02 \cdot (1 - 0.02), 0.02 \cdot (1 + 0.02)]$ around the nominal values 0.01, 0.02 of the coefficients, which is exactly what was stated in section 1.1.2.

Remark 1.2.2. If the perturbation set $\mathcal{Z}$ in (1.2.1) itself is represented as the image of another set $\widehat{\mathcal{Z}}$ under affine mapping $\xi \mapsto \zeta = p + P\xi$, then we can pass from perturbations ζ to perturbations ξ:

$$\begin{aligned} \mathcal{U} &= \left\{\left[\begin{array}{c|c} c^T & d \\ \hline A & b \end{array}\right] = D_0 + \sum_{\ell=1}^{L} \zeta_\ell D_\ell : \zeta \in \mathcal{Z}\right\} \\ &= \left\{\left[\begin{array}{c|c} c^T & d \\ \hline A & b \end{array}\right] = D_0 + \sum_{\ell=1}^{L} [p_\ell + \sum_{k=1}^{K} P_{\ell k}\xi_k] D_\ell : \xi \in \widehat{\mathcal{Z}}\right\} \\ &= \left\{\left[\begin{array}{c|c} c^T & d \\ \hline A & b \end{array}\right] = \underbrace{\left[D_0 + \sum_{\ell=1}^{L} p_\ell D_\ell\right]}_{\widehat{D}_0} + \sum_{k=1}^{K} \xi_k \underbrace{\left[\sum_{\ell=1}^{L} P_{\ell k} D_\ell\right]}_{\widehat{D}_k} : \xi \in \widehat{\mathcal{Z}}\right\}. \end{aligned}$$

It follows that when speaking about perturbation sets with simple geometry (parallelotopes, ellipsoids, etc.), we can normalize these sets to be "standard." For example, a parallelotope is by definition an affine image of a unit box $\{\xi \in \mathbb{R}^k : -1 \leq \xi_j \leq 1, j = 1, ..., k\}$, which gives us the possibility to work with the unit box instead of a general parallelotope. Similarly, an ellipsoid is by definition the image of a unit Euclidean ball $\{\xi \in \mathbb{R}^k : \|x\|_2^2 \equiv x^T x \leq 1\}$ under affine mapping, so that we can work with the standard ball instead of the ellipsoid, etc. We will use this normalization whenever possible.

Note that a *family* of optimization problems like ($\mathrm{LO}_{\mathcal{U}}$), in contrast to a single optimization problem, is not associated by itself with the concepts of feasible/optimal solution and optimal value. How to define these concepts depends of course on the underlying "decision environment." Here we focus on an environment with the following characteristics:

A.1. All decision variables in ($\mathrm{LO}_{\mathcal{U}}$) represent "here and now" decisions; they should be assigned specific numerical values as a result of solving the problem *before* the actual data "reveals itself."

A.2. The decision maker is fully responsible for consequences of the decisions to be made when, and only when, the actual data is within the prespecified uncertainty set $\mathcal{U}$ given by (1.2.1).

A.3. The constraints in ($\mathrm{LO}_{\mathcal{U}}$) are "hard" — we cannot tolerate violations of constraints, even small ones, when the data is in $\mathcal{U}$.

The above assumptions determine, in a more or less unique fashion, what are the meaningful feasible solutions to the uncertain problem ($\mathrm{LO}_{\mathcal{U}}$). By A.1, these should be fixed vectors; by A.2 and A.3, they should be *robust feasible*, that is, they should satisfy all the constraints, whatever the realization of the data from the uncertainty set. We have arrived at the following definition.

Definition 1.2.3. A vector $x \in \mathbb{R}^n$ is a robust feasible solution to ($\mathrm{LO}_{\mathcal{U}}$), if it satisfies all realizations of the constraints from the uncertainty set, that is,

$$Ax \le b \quad \forall (c, d, A, b) \in \mathcal{U}. \tag{1.2.2}$$

As for the objective value to be associated with a meaningful (i.e., robust feasible) solution, assumptions A.1 — A.3 do not prescribe it in a unique fashion. However, "the spirit" of these worst-case-oriented assumptions leads naturally to the following definition:

Definition 1.2.4. Given a candidate solution x, the robust value $\widehat{c}(x)$ of the objective in ($\mathrm{LO}_{\mathcal{U}}$) at x is the largest value of the "true" objective $c^T x + d$ over all realizations of the data from the uncertainty set:

$$\widehat{c}(x) = \sup_{(c,d,A,b)\in\mathcal{U}} [c^T x + d]. \tag{1.2.3}$$

After we agree what are meaningful candidate solutions to the uncertain problem ($\mathrm{LO}_{\mathcal{U}}$) and how to quantify their quality, we can seek the best robust value of the objective among all robust feasible solutions to the problem. This brings us to the central concept of this book, *Robust Counterpart* of an uncertain optimization problem, which is defined as follows:

Definition 1.2.5. The Robust Counterpart of the uncertain LO problem ($\mathrm{LO}_{\mathcal{U}}$) is the optimization problem

$$\min_x \left\{ \widehat{c}(x) = \sup_{(c,d,A,b)\in\mathcal{U}} [c^T x + d] : Ax \le b \ \forall (c, d, A, b) \in \mathcal{U} \right\} \tag{1.2.4}$$

of minimizing the robust value of the objective over all robust feasible solutions to the uncertain problem.

An optimal solution to the Robust Counterpart is called a robust optimal solution to ($\mathrm{LO}_{\mathcal{U}}$), and the optimal value of the Robust Counterpart is called the robust optimal value of ($\mathrm{LO}_{\mathcal{U}}$).

In a nutshell, the robust optimal solution is simply "the best uncertainty-immunized" solution we can associate with our uncertain problem.

Example 1.1.1 continued. Let us find the robust optimal solution to the uncertain problem (Drug). There is exactly one uncertainty-affected "block" in the data, namely, the coefficients of *RawI*, *RawII* in the balance constraint. A candidate solution is thus robust feasible if and only if it satisfies all constraints of (Drug), except for the balance constraint, *and* it satisfies the "worst" realization of the balance constraint. Since *RawI*, *RawII* are nonnegative, the worst realization of the balance constraint is the one where the uncertain coefficients *anI*, *anII* are set to their minimal values in the uncertainty set (these values are 0.00995 and 0.0196, respectively). Since the objective is not affected by the uncertainty, the robust objective values are the same as the original ones. Thus, the RC (Robust Counterpart) of our uncertain problem is the LO problem

$$
\begin{array}{l}
\text{RC(Drug):}\\
\text{RobOpt} = \min\{-100\cdot RawI - 199.9\cdot RawII + 5500\cdot DrugI + 6100\cdot DrugII\}\\
\text{ subject to}\\
\begin{array}{rcl}
0.00995\cdot RawI + 0.0196\cdot RawII - 0.500\cdot DrugI - 0.600\cdot DrugII & \geq & 0\\
RawI + RawII & \leq & 1000\\
90.0\cdot DrugI + 100.0\cdot DrugII & \leq & 2000\\
40.0\cdot DrugI + 50.0\cdot DrugII & \leq & 800\\
100.0\cdot RawI + 199.90\cdot RawII + 700\cdot DrugI + 800\cdot DrugII & \leq & 100000\\
RawI, RawII, DrugI, DrugII & \geq & 0
\end{array}
\end{array}
$$

Solving this problem, we get

$$
\text{RobOpt} = -8294.567; RawI = 877.732, RawII = 0, DrugI = 17.467, DrugII = 0.
$$

The "price" of robustness is the reduction in the promised profit from its nominal optimal value 8819.658 to its robust optimal value 8294.567, that is, by 5.954%. This is much less than the 21% reduction of the actual profit to 6,929 which we may suffer when sticking to the nominal optimal solution when the "true" data are "against" it. Note also that the structure of the robust optimal solution is quite different from the one of the nominal optimal solution: with the robust solution, we shall buy only raw materials RawI, while with the nominal one, only raw materials RawII. The explanation is clear: with the nominal data, RawII as compared to RawI results in a bit smaller per unit price of the active agent (9,995 \$/g vs. 10,000 \$/g). This is why it does not make sense to use RawI with the nominal data. The robust optimal solution takes into account that the uncertainty in *anI* (i.e., the variability of contents of active agent in RawI) is 4 times smaller than that of *anII* (0.5% vs. 2%), which ultimately makes it better to use RawI.

1.2.1 More on Robust Counterparts

We start with several useful observations.

A. The Robust Counterpart (1.2.4) of $\text{LO}_{\mathcal{U}}$ can be rewritten equivalently as the problem

$$
\min_{x,t}\left\{t : \begin{array}{rcl} c^T x - t & \leq & -d \\ Ax & \leq & b \end{array} \right\} \forall (c,d,A,b)\in\mathcal{U}\right\}. \tag{1.2.5}
$$

Note that we can arrive at this problem in another fashion: we first introduce the extra variable t and rewrite instances of our uncertain problem (LO$_\mathcal{U}$) equivalently as

$$\min_{x,t}\left\{t:\begin{array}{rcl} c^Tx-t & \leq & -d \\ Ax & \leq & b \end{array}\right\},$$

thus arriving at an equivalent to (LO$_\mathcal{U}$) uncertain problem in variables x,t with the objective t that is not affected by uncertainty at all. The RC of the reformulated problem is exactly (1.2.5). We see that

> *An uncertain LO problem can always be reformulated as an uncertain LO problem with certain objective. The Robust Counterpart of the reformulated problem has the same objective as this problem and is equivalent to the RC of the original uncertain problem.*

As a consequence, *we lose nothing when restricting ourselves with uncertain LO programs with certain objectives* and we shall frequently use this option in the future.

> We see now why the constant term d in the objective of (1.1.1) should not be neglected, or, more exactly, should not be neglected if it is uncertain. When d is certain, we can account for it by the shift $t \mapsto t-d$ in the slack variable t which affects only the optimal value, but not the optimal solution to the Robust Counterpart (1.2.5). When d is uncertain, there is no "universal" way to eliminate d without affecting the optimal solution to the Robust Counterpart (where d plays the same role as the right hand sides of the original constraints).

B. Assuming that (LO$_\mathcal{U}$) is with certain objective, the Robust Counterpart of the problem is

$$\min_x \left\{c^Tx+d : Ax \leq b,\ \forall (A,b)\in\mathcal{U}\right\} \tag{1.2.6}$$

(note that the uncertainty set is now a set in the space of the constraint data $[A,b]$). We see that

> *The Robust Counterpart of an uncertain LO problem with a certain objective is a purely "constraint-wise" construction: to get RC, we act as follows:*
>
> - *preserve the original certain objective as it is, and*
> - *replace every one of the original constraints*
>
> $$(Ax)_i \leq b_i \Leftrightarrow a_i^Tx \leq b_i \tag{C$_i$}$$
>
> *(a_i^T is i-th row in A) with its Robust Counterpart*
>
> $$a_i^Tx\leq b_i\ \forall [a_i;b_i]\in\mathcal{U}_i, \tag{RC(C$_i$)}$$
>
> *where $\mathcal{U}_i$ is the projection of $\mathcal{U}$ on the space of data of i-th constraint:*
>
> $$\mathcal{U}_i=\{[a_i;b_i] : [A,b]\in\mathcal{U}\}.$$

In particular,

> The RC of an uncertain LO problem with a certain objective remains intact when the original uncertainty set $\mathcal{U}$ is extended to the direct product
> $$\widehat{\mathcal{U}} = \mathcal{U}_1 \times ... \times \mathcal{U}_m$$
> of its projections onto the spaces of data of respective constraints.

Example 1.2.6. The RC of the system of uncertain constraints

$$\{x_1 \geq \zeta_1,\ x_2 \geq \zeta_2\} \tag{1.2.7}$$

with $\zeta \in \mathcal{U} := \{\zeta_1 + \zeta_2 \leq 1, \zeta_1, \zeta_2 \geq 0\}$ is the infinite system of constraints

$$x_1 \geq \zeta_1,\ x_1 \geq \zeta_2\ \ \forall \zeta \in \mathcal{U};$$

on variables x_1, x_2. The latter system is clearly equivalent to the pair of constraints

$$x_1 \geq \max_{\zeta \in \mathcal{U}} \zeta_1 = 1,\ x_2 \geq \max_{\zeta \in \mathcal{U}} \zeta_2 = 1. \tag{1.2.8}$$

The projections of $\mathcal{U}$ to the spaces of data of the two uncertain constraints (1.2.7) are the segments $\mathcal{U}_1 = \{\zeta_1 : 0 \leq \zeta_1 \leq 1\}$, $\mathcal{U}_2 = \{\zeta_2 : 0 \leq \zeta_2 \leq 1\}$, and the RC of (1.2.7) w.r.t.[2] the uncertainty set $\widehat{\mathcal{U}} = \mathcal{U}_1 \times \mathcal{U}_2 = \{\zeta \in \mathbb{R}^2 : 0 \leq \zeta_1, \zeta_2 \leq 1\}$ clearly is (1.2.8).

> The conclusion we have arrived at seems to be counter-intuitive: it says that it is immaterial whether the perturbations of data in different constraints are or are not linked to each other, while intuition says that such a link should be important. We shall see later (chapter 14) that this intuition is valid when a more advanced concept of *Adjustable* Robust Counterpart is considered.

C. If x is a robust feasible solution of (C_i), then x remains robust feasible when we extend the uncertainty set $\mathcal{U}_i$ to its convex hull $\mathrm{Conv}(\mathcal{U}_i)$. Indeed, if $[\bar{a}_i; \bar{b}_i] \in \mathrm{Conv}(\mathcal{U}_i)$, then

$$[\bar{a}_i; \bar{b}_i] = \sum_{j=1}^{J} \lambda_j [a_i^j; b_i^j],$$

with appropriately chosen $[a_i^j; b_i^j] \in \mathcal{U}_i$, $\lambda_j \geq 0$ such that $\sum_j \lambda_j = 1$. We now have

$$\bar{a}_i^T x = \sum_{j=1}^{J} \lambda_j [a_i^j]^T x \leq \sum_j \lambda_j b_i^j = \bar{b}_i,$$

where the inequality is given by the fact that x is feasible for $\mathrm{RC}(\mathrm{C}_i)$ and $[a_i^j; b_i^j] \in \mathcal{U}_i$. We see that $\bar{a}_i^T x \leq \bar{b}_i$ for all $[\bar{a}_i; \bar{b}_i] \in \mathrm{Conv}(\mathcal{U}_i)$, QED.

By similar reasons, the set of robust feasible solutions to (C_i) remains intact when we extend $\mathcal{U}_i$ to the closure of this set. Combining these observations with **B.**, we arrive at the following conclusion:

---

[2] abbr. for "with respect to"

The Robust Counterpart of an uncertain LO problem with a certain objective remains intact when we extend the sets $\mathcal{U}_i$ of uncertain data of respective constraints to their closed convex hulls, and extend $\mathcal{U}$ to the direct product of the resulting sets.

In other words, we lose nothing when assuming from the very beginning that the sets $\mathcal{U}_i$ of uncertain data of the constraints are closed and convex, and $\mathcal{U}$ is the direct product of these sets.

In terms of the parameterization (1.2.1) of the uncertainty sets, the latter conclusion means that

When speaking about the Robust Counterpart of an uncertain LO problem with a certain objective, we lose nothing when assuming that the set $\mathcal{U}_i$ of uncertain data of i-th constraint is given as

$$\mathcal{U}_i = \left\{ [a_i; b_i] = [a_i^0; b_i^0] + \sum_{\ell=1}^{L_i} \zeta_\ell [a_i^\ell; b_i^\ell] : \zeta \in \mathcal{Z}_i \right\}, \tag{1.2.9}$$

with a closed and convex perturbation set $\mathcal{Z}_i$.

D. An important modeling issue. In the usual — with certain data — Linear Optimization, constraints can be modeled in various equivalent forms. For example, we can write:

$$\begin{array}{ll} (a) & a_1x_1 + a_2x_2 \leq a_3 \\ (b) & a_4x_1 + a_5x_2 = a_6 \\ (c) & x_1 \geq 0, x_2 \geq 0 \end{array} \tag{1.2.10}$$

or, equivalently,

$$\begin{array}{ll} (a) & a_1x_1 + a_2x_2 \leq a_3 \\ (b.1) & a_4x_1 + a_5x_2 \leq a_6 \\ (b.2) & -a_5x_1 - a_5x_2 \leq -a_6 \\ (c) & x_1 \geq 0, x_2 \geq 0. \end{array} \tag{1.2.11}$$

Or, equivalently, by adding a slack variable s,

$$\begin{array}{ll} (a) & a_1x_1 + a_2x_2 + s = a_3 \\ (b) & a_4x_1 + a_5x_2 = a_6 \\ (c) & x_1 \geq 0, x_2 \geq 0, s \geq 0. \end{array} \tag{1.2.12}$$

However, when (part of) the data $a_1, ..., a_6$ become *uncertain*, not all of these equivalences remain valid: the RCs of our now uncertainty-affected systems of constraints are not equivalent to each other. Indeed, denoting the uncertainty set by $\mathcal{U}$, the RCs read, respectively,

$$\left.\begin{array}{ll} (a) & a_1x_1 + a_2x_2 \leq a_3 \\ (b) & a_4x_1 + a_5x_2 = a_6 \\ (c) & x_1 \geq 0, x_2 \geq 0 \end{array}\right\} \forall a = [a_1; ...; a_6] \in \mathcal{U}. \tag{1.2.13}$$

$$\left.\begin{array}{ll}(a) & a_1x_1 + a_2x_2 \leq a_3\\ (b.1) & a_4x_1 + a_5x_2 \leq a_6\\ (b.2) & -a_5x_1 - a_5x_2 \leq -a_6\\ (c) & x_1 \geq 0, x_2 \geq 0\end{array}\right\}\ \forall a = [a_1;...;a_6] \in \mathcal{U}. \tag{1.2.14}$$

$$\left.\begin{array}{ll}(a) & a_1x_1 + a_2x_2 + s = a_3\\ (b) & a_4x_1 + a_5x_2 = a_6\\ (c) & x_1 \geq 0, x_2 \geq 0, s \geq 0\end{array}\right\}\ \forall a = [a_1;...;a_6] \in \mathcal{U}. \tag{1.2.15}$$

It is immediately seen that while the first and the second RCs are equivalent to each other,[3] they are *not* equivalent to the third RC. The latter RC is more conservative than the first two, meaning that whenever (x_1, x_2) can be extended, by a properly chosen s, to a feasible solution of (1.2.15), (x_1, x_2) is feasible for (1.2.13)≡(1.2.14) (this is evident), but not necessarily vice versa. In fact, the gap between (1.2.15) and (1.2.13)≡(1.2.14) can be quite large. To illustrate the latter claim, consider the case where the uncertainty set is

$$\mathcal{U} = \{a = a_\zeta := [1+\zeta; 2+\zeta; 4-\zeta; 4+\zeta; 5-\zeta; 9] : -\rho \leq \zeta \leq \rho\},$$

where ζ is the data perturbation. In this situation, $x_1 = 1$, $x_2 = 1$ is a feasible solution to (1.2.13)≡(1.2.14), provided that the uncertainty level ρ is $\leq 1/3$:

$$(1+\zeta)\cdot 1 + (2+\zeta)\cdot 1 \leq 4-\zeta\,\forall(\zeta : |\zeta| \leq \rho \leq 1/3)\ \&(4+\zeta)\cdot 1 + (5-\zeta)\cdot 1 = 9\ \forall\zeta.$$

At the same time, when $\rho > 0$, our solution $(x_1 = 1, x_2 = 1)$ cannot be extended to a feasible solution of (1.2.15), since the latter system of constraints is infeasible and remains so even after eliminating the equality (1.2.15.b).

Indeed, in order for x_1, x_2, s to satisfy (1.2.15.a) for all $a \in \mathcal{U}$, we should have

$$x_1 + 2x_2 + s + \zeta[x_1 + x_2] = 4 - \zeta\ \forall(\zeta : |\zeta| \leq \rho);$$

when $\rho > 0$, we therefore should have $x_1 + x_2 = -1$, which contradicts (1.2.15.c)

The origin of the outlined phenomenon is clear. Evidently the inequality $a_1x_1 + a_2x_2 \leq a_3$, where all a_i and x_i are fixed reals, holds true if and only if we can "certify" the inequality by pointing out a real $s \geq 0$ such that $a_1x_1 + a_2x_2 + s = a_3$. When the data a_1, a_2, a_3 become uncertain, the restriction on (x_1, x_2) to be robust feasible for the uncertain inequality $a_1x_1 + a_2x_2 \leq a_3$ for all $a \in \mathcal{U}$ reads, "in terms of certificate," as

$$\forall a \in \mathcal{U}\, \exists s \geq 0 : a_1x_1 + a_2x_2 + s = a_3,$$

that is, the certificate s should be allowed to depend on the true data. In contrast to this, in (1.2.15) we require from both the decision variables x *and the slack variable ("the certificate")* s to be independent of the true data, which is by far too conservative.

What can be learned from the above examples is that when modeling an uncertain LO problem one should avoid whenever possible converting inequality

[3] Clearly, this always is the case when an equality constraint, certain or uncertain alike, is replaced with a pair of opposite inequalities.

constraints into equality ones, unless all the data in the constraints in question are certain. Aside from avoiding slack variables,[4] this means that restrictions like "total expenditure cannot exceed the budget," or "supply should be at least the demand," which in LO problems with certain data can harmlessly be modeled by equalities, in the case of uncertain data should be modeled by inequalities. This is in full accordance with common sense saying, e.g., that when the demand is uncertain and its satisfaction is a must, it would be unwise to forbid surplus in supply. Sometimes a good for the RO methodology modeling requires eliminating "state variables" — those which are readily given by variables representing actual decisions — via the corresponding "state equations." For example, time dynamics of an inventory is given in the simplest case by the state equations

$$\begin{aligned} &x_0 = c \\ &x_{t+1} = x_t + q_t - d_t,\ t = 0, 1, ..., T, \end{aligned}$$

where x_t is the inventory level at time t, d_t is the (uncertain) demand in period $[t, t+1)$, and variables q_t represent actual decisions – replenishment orders at instants $t = 0, 1, ..., T$. A wise approach to the RO processing of such an inventory problem would be to eliminate the state variables x_t by setting

$$x_t = c + \sum_{\tau=1}^{t-1} q_\tau,\ t = 0, 1, 2, ..., T + 1,$$

and to get rid of the state equations. As a result, typical restrictions on state variables (like "x_t should stay within given bounds" or "total holding cost should not exceed a given bound") will become uncertainty-affected inequality constraints on the actual decisions q_t, and we can process the resulting inequality-constrained uncertain LO problem via its RC.[5]

1.2.2 What is Ahead

After introducing the concept of the Robust Counterpart of an uncertain LO problem, we confront two major questions:

i) What is the "computational status" of the RC? When is it possible to process the RC efficiently?

ii) How to come-up with meaningful uncertainty sets?

The first of these questions, to be addressed in depth in section 1.3, is a "structural" one: what should be the structure of the uncertainty set in order to make the RC computationally tractable? Note that the RC as given by (1.2.5) or (1.2.6) is a *semi-infinite* LO program, that is, an optimization program with simple linear

---

[4] Note that slack variables do not represent actual decisions; thus, their presence in an LO model contradicts assumption A.1, and thus can lead to too conservative, or even infeasible, RCs.

[5] For more advanced robust modeling of uncertainty-affected multi-stage inventory, see chapter 14.

objective and *infinitely many* linear constraints. In principle, such a problem can be "computationally intractable" — NP-hard.

Example 1.2.7. Consider an uncertain "essentially linear" constraint

$$\{\|Px - p\|_1 \le 1\}_{[P;p]\in\mathcal{U}}, \tag{1.2.16}$$

where $\|z\|_1 = \sum_j |z_j|$, and assume that the matrix P is certain, while the vector p is uncertain and is parameterized by perturbations from the unit box:

$$p \in \{p = B\zeta : \|\zeta\|_\infty \le 1\},$$

where $\|\zeta\|_\infty = \max_\ell |\zeta_\ell|$ and B is a given positive semidefinite matrix. To check whether $x = 0$ is robust feasible is exactly the same as to verify whether $\|B\zeta\|_1 \le 1$ whenever $\|\zeta\|_\infty \le 1$; or, due to the evident relation $\|u\|_1 = \max_{\|\eta\|_\infty \le 1} \eta^T u$, the same as to check whether $\max_{\eta,\zeta} \left\{\eta^T B\zeta : \|\eta\|_\infty \le 1, \|\zeta\|_\infty \le 1\right\} \le 1$. The maximum of the bilinear form $\eta^T B\zeta$ with positive semidefinite B over η, ζ varying in a convex symmetric neighborhood of the origin is always achieved when $\eta = \zeta$ (you may check this by using the polarization identity $\eta^T B\zeta = \frac{1}{4}(\eta+\zeta)^T B(\eta+\zeta) - \frac{1}{4}(\eta-\zeta)^T B(\eta-\zeta)$). Thus, to check whether $x = 0$ is robust feasible for (1.2.16) is the same as to check whether the maximum of a given nonnegative quadratic form $\zeta^T B\zeta$ over the unit box is ≤ 1. The latter problem is known to be NP-hard,[6] and therefore so is the problem of checking robust feasibility for (1.2.16).

The second of the above is a modeling question, and as such, goes beyond the scope of purely theoretical considerations. However, theory, as we shall see in section 2.1, contributes significantly to this modeling issue.

1.3 TRACTABILITY OF ROBUST COUNTERPARTS

In this section, we investigate the "computational status" of the RC of uncertain LO problem. The situation here turns out to be as good as it could be: we shall see, essentially, that *the RC of the uncertain LO problem with uncertainty set $\mathcal{U}$ is computationally tractable whenever the convex uncertainty set $\mathcal{U}$ itself is computationally tractable.* The latter means that we know in advance the affine hull of $\mathcal{U}$, a point from the relative interior of $\mathcal{U}$, and we have access to an efficient *membership oracle* that, given on input a point u, reports whether $u \in \mathcal{U}$. This can be reformulated as a precise mathematical statement; however, we will prove a slightly restricted version of this statement that does not require long excursions into complexity theory.

1.3.1 The Strategy

Our strategy will be as follows. First, we restrict ourselves to uncertain LO problems with a certain objective — we remember from item **A** in Section 1.2.1 that we lose

[6] In fact, it is NP-hard to compute the maximum of a nonnegative quadratic form over the unit box with inaccuracy less than 4% [61].

nothing by this restriction. Second, all we need is a "computationally tractable" representation of the RC of a *single* uncertain linear constraint, that is, an equivalent representation of the RC by an explicit (and "short") system of efficiently verifiable convex inequalities. Given such representations for the RCs of every one of the constraints of our uncertain problem and putting them together (cf. item **B** in Section 1.2.1), we reformulate the RC of the problem as the problem of minimizing the original linear objective under a finite (and short) system of explicit convex constraints, and thus — as a computationally tractable problem.

To proceed, we should explain first what does it mean to represent a constraint by a system of convex inequalities. Everyone understands that the system of 4 constraints on 2 variables,

$$x_1 + x_2 \leq 1, x_1 - x_2 \leq 1, -x_1 + x_2 \leq 1, -x_1 - x_2 \leq 1, \tag{1.3.1}$$

represents the nonlinear inequality

$$|x_1| + |x_2| \leq 1 \tag{1.3.2}$$

in the sense that both (1.3.2) and (1.3.1) define the same feasible set. Well, what about the claim that the system of 5 linear inequalities

$$-u_1 \leq x_1 \leq u_1, -u_2 \leq x_2 \leq u_2, u_1 + u_2 \leq 1 \tag{1.3.3}$$

represents the same set as (1.3.2)? Here again everyone will agree with the claim, although we cannot justify the claim in the former fashion, since the feasible sets of (1.3.2) and (1.3.3) live in different spaces and therefore cannot be equal to each other!

What actually is meant when speaking about "equivalent representations of problems/constraints" in Optimization can be formalized as follows:

Definition 1.3.1. A set $X^+ \subset \mathbb{R}^n_x \times \mathbb{R}^k_u$ is said to represent a set $X \subset \mathbb{R}^n_x$, if the projection of X^+ onto the space of x-variables is exactly X, i.e., $x \in X$ if and only if there exists $u \in \mathbb{R}^k_u$ such that $(x, u) \in X^+$:

$$X = \left\{x : \exists u : (x, u) \in X^+\right\}.$$

A system of constraints $\mathcal{S}^+$ in variables $x \in \mathbb{R}^n_x$, $u \in \mathbb{R}^k_u$ is said to represent a system of constraints $\mathcal{S}$ in variables $x \in \mathbb{R}^n_x$, if the feasible set of the former system represents the feasible set of the latter one.

With this definition, it is clear that the system (1.3.3) indeed represents the constraint (1.3.2), and, more generally, that the system of $2n + 1$ linear inequalities

$$-u_j \leq x_j \leq u_j,\, j = 1, ..., n, \sum_j u_j \leq 1$$

in variables x, u represents the constraint

$$\sum_j |x_j| \leq 1.$$

To understand how powerful this representation is, note that to represent the same constraint in the style of (1.3.1), that is, without extra variables, it would take as much as 2^n linear inequalities.

Coming back to the general case, assume that we are given an optimization problem

$$\min_x \{f(x) \text{ s.t. } x \text{ satisfies } \mathcal{S}_i,\ i = 1, ..., m\}, \tag{P}$$

where $\mathcal{S}_i$ are systems of constraints in variables x, and that we have in our disposal systems $\mathcal{S}_i^+$ of constraints in variables x, v^i which represent the systems $\mathcal{S}_i$. Clearly, the problem

$$\min_{x,v^1,...,v^m} \left\{f(x) \text{ s.t. } (x, v^i) \text{ satisfies } \mathcal{S}_i^+,\ i = 1, ..., m\right\} \tag{P$^+$}$$

is equivalent to (P): the x component of every feasible solution to (P$^+$) is feasible for (P) with the same value of the objective, and the optimal values in the problems are equal to each other, so that the x component of an ϵ-optimal (in terms of the objective) feasible solution to (P$^+$) is an ϵ-optimal feasible solution to (P). We shall say that (P$^+$) represents equivalently the original problem (P). What is important here, is that a representation can possess desired properties that are absent in the original problem. For example, an appropriate representation can convert the problem of the form $\min\limits_x\{\|Px-p\|_1 : Ax \leq b\}$ with n variables, m linear constraints, and k-dimensional vector p, into an LO problem with $n+k$ variables and $m+2k+1$ linear inequality constraints, etc. Our goal now is to build a representation capable of expressing equivalently a semi-infinite linear constraint (specifically, the robust counterpart of an uncertain linear inequality) as a finite system of explicit convex constraints, with the ultimate goal to use these representations in order to convert the RC of an uncertain LO problem into an explicit (and as such, computationally tractable) convex program.

The outlined strategy allows us to focus on a *single* uncertainty-affected linear inequality — a family

$$\left\{a^T x \leq b\right\}_{[a;b]\in\mathcal{U}}, \tag{1.3.4}$$

of linear inequalities with the data varying in the uncertainty set

$$\mathcal{U} = \left\{[a; b] = [a^0; b^0] + \sum_{\ell=1}^{L} \zeta_\ell [a^\ell; b^\ell] : \zeta \in \mathcal{Z}\right\} \tag{1.3.5}$$

— and on "tractable representation" of the RC

$$a^T x \leq b \quad \forall \left([a; b] = [a^0; b^0] + \sum_{\ell=1}^{L} \zeta_\ell [a^\ell; b^\ell] : \zeta \in \mathcal{Z}\right) \tag{1.3.6}$$

of this uncertain inequality.

By reasons indicated in item **C** of Section 1.2.1, we assume from now on that the associated perturbation set $\mathcal{Z}$ is convex.

1.3.2 Tractable Representation of (1.3.6): Simple Cases

We start with the cases where the desired representation can be found by "bare hands," specifically, the cases of *interval* and *simple ellipsoidal* uncertainty.

Example 1.3.2. Consider the case of *interval uncertainty*, where $\mathcal{Z}$ in (1.3.6) is a box. W.l.o.g.[7] we can normalize the situation by assuming that

$$\mathcal{Z} = \mathrm{Box}_1 \equiv \{\zeta \in \mathbb{R}^L : \|\zeta\|_\infty \le 1\}.$$

In this case, (1.3.6) reads

$$\begin{array}{lll} & [a^0]^T x + \sum_{\ell=1}^L \zeta_\ell [a^\ell]^T x \le b^0 + \sum_{\ell=1}^L \zeta_\ell b^\ell & \forall(\zeta : \|\zeta\|_\infty \le 1) \\ \Leftrightarrow & \sum_{\ell=1}^L \zeta_\ell [[a^\ell]^T x - b^\ell] \le b^0 - [a^0]^T x & \forall(\zeta : |\zeta_\ell| \le 1,\ \ell = 1, ..., L) \\ \Leftrightarrow & \max_{-1 \le \zeta_\ell \le 1} \left[\sum_{\ell=1}^L \zeta_\ell [[a^\ell]^T x - b^\ell] \right] \le b^0 - [a^0]^T x & \end{array}$$

The concluding maximum in the chain is clearly $\sum_{\ell=1}^L |[a^\ell]^T x - b^\ell|$, and we arrive at the representation of (1.3.6) by the explicit convex constraint

$$[a^0]^T x + \sum_{\ell=1}^L |[a^\ell]^T x - b^\ell| \le b^0, \tag{1.3.7}$$

which in turn admits a representation by a system of linear inequalities:

$$\left\{ \begin{array}{l} -u_\ell \le [a^\ell]^T x - b^\ell \le u_\ell,\ \ell = 1, ..., L, \\ [a^0]^T x + \sum_{\ell=1}^L u_\ell \le b^0. \end{array} \right. \tag{1.3.8}$$

Example 1.3.3. Consider the case of *ellipsoidal uncertainty* where $\mathcal{Z}$ in (1.3.6) is an ellipsoid. W.l.o.g. we can normalize the situation by assuming that $\mathcal{Z}$ is merely the ball of radius Ω centered at the origin:

$$\mathcal{Z} = \mathrm{Ball}_\Omega = \{\zeta \in \mathbb{R}^L : \|\zeta\|_2 \le \Omega\}.$$

In this case, (1.3.6) reads

$$\begin{array}{lll} & [a^0]^T x + \sum_{\ell=1}^L \zeta_\ell [a^\ell]^T x \le b^0 + \sum_{\ell=1}^L \zeta_\ell b^\ell & \forall(\zeta : \|\zeta\|_2 \le \Omega) \\ \Leftrightarrow & \max_{\|\zeta\|_2 \le \Omega} \left[\sum_{\ell=1}^L \zeta_\ell [[a^\ell]^T x - b^\ell] \right] \le b^0 - [a^0]^T x & \\ \Leftrightarrow & \Omega \sqrt{\sum_{\ell=1}^L ([a^\ell]^T x - b^\ell)^2} \le b^0 - [a^0]^T x, & \end{array}$$

and we arrive at the representation of (1.3.6) by the explicit convex constraint ("conic quadratic inequality")

$$[a^0]^T x + \Omega \sqrt{\sum_{\ell=1}^L ([a^\ell]^T x - b^\ell)^2} \le b^0. \tag{1.3.9}$$

[7] abbr. for "without loss of generality."

1.3.3 Tractable Representation of (1.3.6): General Case

Now consider a rather general case when the perturbation set $\mathcal{Z}$ in (1.3.6) is given by a *conic representation* (cf. section A.2.4 in Appendix):

$$\mathcal{Z} = \left\{\zeta \in \mathbb{R}^L : \exists u \in \mathbb{R}^K : P\zeta + Qu + p \in \mathbf{K}\right\}, \tag{1.3.10}$$

where $\mathbf{K}$ is a closed convex pointed cone in $\mathbb{R}^N$ with a nonempty interior, P, Q are given matrices and p is a given vector. In the case when $\mathbf{K}$ is *not* a polyhedral cone, assume that this representation is strictly feasible:

$$\exists(\bar{\zeta}, \bar{u}) : P\bar{\zeta} + Q\bar{u} + p \in \mathrm{int}K. \tag{1.3.11}$$

Theorem 1.3.4. Let the perturbation set $\mathcal{Z}$ be given by (1.3.10), and in the case of non-polyhedral $\mathbf{K}$, let also (1.3.11) take place. Then the semi-infinite constraint (1.3.6) can be represented by the following system of conic inequalities in variables $x \in \mathbb{R}^n, y \in \mathbb{R}^N$:

$$\begin{array}{l} p^T y + [a^0]^T x \le b^0, \\ Q^T y = 0, \\ (P^T y)_\ell + [a^\ell]^T x = b^\ell, \ell = 1, ..., L, \\ y \in \mathbf{K}_*, \end{array} \tag{1.3.12}$$

where $\mathbf{K}_* = \{y : y^T z \ge 0 \,\forall z \in \mathbf{K}\}$ is the cone dual to $\mathbf{K}$.

Proof. We have

$$\begin{array}{rc} & x \text{ is feasible for (1.3.6)} \\ \Leftrightarrow & \sup\limits_{\zeta \in \mathcal{Z}} \Big\{ \underbrace{[a^0]^T x - b^0}_{d[x]} + \sum\limits_{\ell=1}^{L} \zeta_\ell \underbrace{\left[[a^\ell]^T x - b^\ell\right]}_{c_\ell[x]} \Big\} \le 0 \\ \Leftrightarrow & \sup\limits_{\zeta \in \mathcal{Z}} \left\{c^T[x]\zeta + d[x]\right\} \le 0 \\ \Leftrightarrow & \sup\limits_{\zeta \in \mathcal{Z}} c^T[x]\zeta \le -d[x] \\ \Leftrightarrow & \max\limits_{\zeta, v} \left\{c^T[x]\zeta : P\zeta + Qv + p \in \mathbf{K}\right\} \le -d[x]. \end{array}$$

The concluding relation says that x is feasible for (1.3.6) if and only if the optimal value in the conic program

$$\max_{\zeta, v} \left\{c^T[x]\zeta : P\zeta + Qv + p \in \mathbf{K}\right\} \tag{CP}$$

is $\le -d[x]$. Assume, first, that (1.3.11) takes place. Then (CP) is strictly feasible, and therefore, applying the Conic Duality Theorem (Theorem A.2.1), the optimal value in (CP) is $\le -d[x]$ if and only if the optimal value in the conic dual to the (CP) problem

$$\min_y \left\{p^T y : Q^T y = 0, P^T y = -c[x], y \in \mathbf{K}_*\right\}, \tag{CD}$$

is attained and is $\le -d[x]$. Now assume that $\mathbf{K}$ is a polyhedral cone. In this case the usual LO Duality Theorem, (which does not require the validity of (1.3.11)), yields exactly the same conclusion: the optimal value in (CP) is $\le -d[x]$ if and only if the optimal value in (CD) is achieved and is $\le -d[x]$. In other words, under the premise of the Theorem, x is feasible for (1.3.6) if and only if (CD) has a feasible solution y with $p^T y \leq -d[x]$. □

Observing that nonnegative orthants, Lorentz and Semidefinite cones are self-dual, we derive from Theorem 1.3.4 the following corollary:

Corollary 1.3.5. Let the nonempty perturbation set in (1.3.6) be:

(i) polyhedral, i.e., given by (1.3.10) with a nonnegative orthant $\mathbb{R}^N_+$ in the role of $\mathbf{K}$, or

(ii) conic quadratic representable, i.e., given by (1.3.10) with a direct product $\mathbf{L}^{k_1} \times ... \times \mathbf{L}^{k_m}$ of Lorentz cones $\mathbf{L}^k = \{x \in \mathbb{R}^k : x_k \geq \sqrt{x_1^2 + ... + x_{k-1}{}^2}\}$ in the role of $\mathbf{K}$, or

(iii) semidefinite representable, i.e., given by (1.3.10) with the positive semidefinite cone $\mathbf{S}^k_+$ in the role of $\mathbf{K}$.

In the cases of (ii), (iii) assume in addition that (1.3.11) holds true. Then the Robust Counterpart (1.3.6) of the uncertain linear inequality (1.3.4) — (1.3.5) with the perturbation set $\mathcal{Z}$ admits equivalent reformulation as an explicit system of

— linear inequalities, in the case of (i),

— conic quadratic inequalities, in the case of (ii),

— linear matrix inequalities, in the case of (iii).

In all cases, the size of the reformulation is polynomial in the number of variables in (1.3.6) and the size of the conic description of $\mathcal{Z}$, while the data of the reformulation is readily given by the data describing, via (1.3.10), the perturbation set $\mathcal{Z}$.

Remark 1.3.6. A. Usually, the cone $\mathbf{K}$ participating in (1.3.10) is the direct product of simpler cones $\mathbf{K}^1, ..., \mathbf{K}^S$, so that representation (1.3.10) takes the form

$$\mathcal{Z} = \{\zeta : \exists u^1, ..., u^S : P_s\zeta + Q_s u^s + p_s \in \mathbf{K}^s,\ s = 1, ..., S\}. \tag{1.3.13}$$

In this case, (1.3.12) becomes the system of conic constraints in variables $x, y^1, ..., y^S$ as follows:

$$\begin{array}{l} \sum\limits_{s=1}^{S} p_s^T y^s + [a^0]^T x \leq b^0, \\ Q_s^T y^s = 0,\ s = 1, ..., S, \\ \sum\limits_{s=1}^{S} (P_s^T y^s)_\ell + [a^\ell]^T x = b^\ell, \ell = 1, ..., L, \\ y^s \in \mathbf{K}_*^s, s = 1, ..., S, \end{array} \tag{1.3.14}$$

where K_*^s is the cone dual to K^s.

B. Uncertainty sets given by LMIs seem "exotic"; however, they can arise under quite realistic circumstances, see section 1.4.

1.3.3.1 Examples

We are about to apply Theorem 1.3.4 to build tractable reformulations of the semi-infinite inequality (1.3.6) in two particular cases. While at a first glance no natural "uncertainty models" lead to the "strange" perturbation sets we are about to consider, it will become clear later that these sets are of significant importance — they allow one to model *random* uncertainty.

Example 1.3.7. *$\mathcal{Z}$ is the intersection of concentric co-axial box and ellipsoid, specifically,*

$$\mathcal{Z} = \{\zeta \in \mathbb{R}^L : -1 \le \zeta_\ell \le 1, \ell \le L, \sqrt{\sum_{\ell=1}^{L} \zeta_\ell^2/\sigma_\ell^2} \le \Omega\}, \tag{1.3.15}$$

where $\sigma_\ell > 0$ and $\Omega > 0$ are given parameters.

Here representation (1.3.13) becomes

$$\mathcal{Z} = \{\zeta \in \mathbb{R}^L : P_1\zeta + p_1 \in \mathbf{K}^1, P_2\zeta + p_2 \in \mathbf{K}^2\},$$

where

- $P_1\zeta \equiv [\zeta; 0]$, $p_1 = [0_{L\times 1}; 1]$ and $\mathbf{K}^1 = \{(z,t) \in \mathbb{R}^L \times \mathbb{R} : t \ge \|z\|_\infty\}$, whence $\mathbf{K}^1_* = \{(z,t) \in \mathbb{R}^L \times \mathbb{R} : t \ge \|z\|_1\}$;
- $P_2\zeta = [\Sigma^{-1}\zeta; 0]$ with $\Sigma = \mathrm{Diag}\{\sigma_1, ..., \sigma_L\}$, $p_2 = [0_{L\times 1}; \Omega]$ and $\mathbf{K}^2$ is the Lorentz cone of the dimension $L+1$ (whence $\mathbf{K}^2_* = \mathbf{K}^2$)

Setting $y^1 = [\eta_1; \tau_1]$, $y^2 = [\eta_2; \tau_2]$ with one-dimensional τ_1, τ_2 and L-dimensional η_1, η_2, (1.3.14) becomes the following system of constraints in variables τ, η, x:

$$\begin{array}{llcl}
(a) & \tau_1 + \Omega\tau_2 + [a^0]^T x & \le & b^0, \\
(b) & (\eta_1 + \Sigma^{-1}\eta_2)_\ell & = & b^\ell - [a^\ell]^T x,\ \ell = 1, ..., L, \\
(c) & \|\eta_1\|_1 & \le & \tau_1 \quad [\Leftrightarrow [\eta_1; \tau_1] \in \mathbf{K}^1_*], \\
(d) & \|\eta_2\|_2 & \le & \tau_2 \quad [\Leftrightarrow [\eta_2; \tau_2] \in \mathbf{K}^2_*].
\end{array}$$

We can eliminate from this system the variables τ_1, τ_2 — for every feasible solution to the system, we have $\tau_1 \ge \bar\tau_1 \equiv \|\eta_1\|_1$, $\tau_2 \ge \bar\tau_2 \equiv \|\eta_2\|_2$, and the solution obtained when replacing τ_1, τ_2 with $\bar\tau_1$, $\bar\tau_2$ still is feasible. The reduced system in variables x, $z = \eta_1$, $w = \Sigma^{-1}\eta_2$ reads

$$\begin{array}{rcl}
\sum_{\ell=1}^{L} |z_\ell| + \Omega\sqrt{\sum_\ell \sigma_\ell^2 w_\ell^2} + [a^0]^T x & \le & b^0, \\
z_\ell + w_\ell & = & b^\ell - [a^\ell]^T x,\ \ell = 1, ..., L,
\end{array} \tag{1.3.16}$$

which is also a representation of (1.3.6), (1.3.15).

Example 1.3.8. [*"budgeted uncertainty"*] Consider the case where $\mathcal{Z}$ is the intersection of $\|\cdot\|_\infty$- and $\|\cdot\|_1$-balls, specifically,

$$\mathcal{Z} = \{\zeta \in \mathbb{R}^L : \|\zeta\|_\infty \le 1,\ \|\zeta\|_1 \le \gamma\}, \tag{1.3.17}$$

where γ, $1 \le \gamma \le L$, is a given "uncertainty budget."

Here representation (1.3.13) becomes

$$\mathcal{Z} = \{\zeta \in \mathbb{R}^L : P_1\zeta + p_1 \in \mathbf{K}^1, P_2\zeta + p_2 \in \mathbf{K}^2\},$$

where

- $P_1\zeta \equiv [\zeta;0]$, $p_1 = [0_{L\times 1};1]$ and $\mathbf{K}^1 = \{[z;t] \in \mathbb{R}^L \times \mathbb{R} : t \geq \|z\|_\infty\}$, whence $\mathbf{K}^1_* = \{[z;t] \in \mathbb{R}^L \times \mathbb{R} : t \geq \|z\|_1\}$;
- $P_2\zeta = [\zeta;0]$, $p_2 = [0_{L\times 1};\gamma]$ and $\mathbf{K}^2 = \mathbf{K}^1_* = \{[z;t] \in \mathbb{R}^L \times \mathbb{R} : t \geq \|z\|_1\}$, whence $\mathbf{K}^2_* = \mathbf{K}^1$.

Setting $y^1 = [z;\tau_1]$, $y^2 = [w;\tau_2]$ with one-dimensional τ and L-dimensional z, w, system (1.3.14) becomes the following system of constraints in variables τ_1, τ_2, z, w, x:

$$\begin{array}{lrcl}
(a) & \tau_1 + \gamma\tau_2 + [a^0]^T x & \leq & b^0, \\
(b) & (z+w)_\ell & = & b^\ell - [a^\ell]^T x,\ \ell = 1,...,L, \\
(c) & \|z\|_1 & \leq & \tau_1 \quad [\Leftrightarrow [\eta_1;\tau_1] \in \mathbf{K}^1_*], \\
(d) & \|w\|_\infty & \leq & \tau_2 \quad [\Leftrightarrow [\eta_2;\tau_2] \in \mathbf{K}^2_*].
\end{array}$$

Same as in Example 1.3.7, we can eliminate the τ-variables, arriving at a representation of (1.3.6), (1.3.17) by the following system of constraints in variables x, z, w:

$$\begin{array}{rcl}
\sum_{\ell=1}^{L} |z_\ell| + \gamma \max_\ell |w_\ell| + [a^0]^T x & \leq & b^0, \\
z_\ell + w_\ell & = & b^\ell - [a^\ell]^T x,\ \ell = 1,...,L,
\end{array} \tag{1.3.18}$$

which can be further converted into the system of linear inequalities in z, w and additional variables.

1.4 NON-AFFINE PERTURBATIONS

In the first reading this section can be skipped.

So far we have assumed that the uncertain data of an uncertain LO problem are *affinely* parameterized by a perturbation vector ζ varying in a closed convex set $\mathcal{Z}$. We have seen that this assumption, combined with the assumption that $\mathcal{Z}$ is computationally tractable, implies tractability of the RC. What happens when the perturbations enter the uncertain data in a nonlinear fashion? Assume w.l.o.g. that every entry a in the uncertain data is of the form

$$a = \sum_{k=1}^{K} c^a_k f_k(\zeta),$$

where c^a_k are given coefficients (depending on the data entry in question) and $f_1(\zeta), ..., f_K(\zeta)$ are certain basic functions, perhaps non-affine, defined on the perturbation set $\mathcal{Z}$. Assuming w.l.o.g. that the objective is certain, we still can define the RC of our uncertain problem as the problem of minimizing the original objective over the set of robust feasible solutions, those which remain feasible for all values of the data coming from $\zeta \in \mathcal{Z}$, but what about the tractability of this RC? An immediate observation is that the case of nonlinearly perturbed data can be immediately reduced to the one where the data are affinely perturbed. To this end, it suffices to pass from the original perturbation vector ζ to the new vector

$$\widehat{\zeta}[\zeta] = [\zeta_1; ...; \zeta_L; f_1(\zeta); ...; f_K(\zeta)].$$

As a result, the uncertain data become *affine* functions of the new perturbation vector $\widehat{\zeta}$ which now runs through the image $\widetilde{\mathcal{Z}} = \widehat{\zeta}[\mathcal{Z}]$ of the original uncertainty set $\mathcal{Z}$ under the mapping $\zeta \mapsto \widehat{\zeta}[\zeta]$. As we know, in the case of affine data perturbations the RC remains intact when replacing a given perturbation set with its closed convex hull. Thus, we can think about our uncertain LO problem as an affinely perturbed problem where the perturbation vector is $\widehat{\zeta}$, and this vector runs through the closed convex set $\widehat{\mathcal{Z}} = \mathrm{cl}\,\mathrm{Conv}(\widehat{\zeta}[\mathcal{Z}])$. We see that formally speaking, the case of general-type perturbations can be reduced to the one of affine perturbations. This, unfortunately, does not mean that non-affine perturbations do not cause difficulties. Indeed, in order to end up with a computationally tractable RC, we need more than affinity of perturbations and convexity of the perturbation set — we need this set to be computationally tractable. And the set $\widehat{\mathcal{Z}} = \mathrm{cl}\,\mathrm{Conv}(\widehat{\zeta}[\mathcal{Z}])$ may fail to satisfy this requirement even when both $\mathcal{Z}$ and the *nonlinear* mapping $\zeta \mapsto \widehat{\zeta}[\zeta]$ are simple, e.g., when $\mathcal{Z}$ is a box and $\widehat{\zeta} = [\zeta; \{\zeta_\ell \zeta_r\}_{\ell,r=1}^L]$, (i.e., when the uncertain data are quadratically perturbed by the original perturbations ζ).

We are about to present two generic cases where the difficulty just outlined does not occur (for justification and more examples, see section 14.3.2).

Ellipsoidal perturbation set $\mathcal{Z}$, quadratic perturbations. Here $\mathcal{Z}$ is an ellipsoid, and the basic functions f_k are the constant, the coordinates of ζ and the pairwise products of these coordinates. This means that the uncertain data entries are quadratic functions of the perturbations. W.l.o.g. we can assume that the ellipsoid $\mathcal{Z}$ is centered at the origin: $\mathcal{Z} = \{\zeta : \|Q\zeta\|_2 \leq 1\}$, where $\mathrm{Ker}Q = \{0\}$. In this case, representing $\widehat{\zeta}[\zeta]$ as the matrix $\left[\begin{array}{c|c} & \zeta^T \\ \hline \zeta & \zeta\zeta^T \end{array}\right]$, we have the following semidefinite representation of $\widehat{\mathcal{Z}} = \mathrm{cl}\,\mathrm{Conv}(\widehat{\zeta}[\mathcal{Z}])$:

$$\widehat{\mathcal{Z}} = \left\{ \left[\begin{array}{c|c} & w^T \\ \hline w & W \end{array}\right] : \left[\begin{array}{c|c} 1 & w^T \\ \hline w & W \end{array}\right] \succeq 0, \mathrm{Tr}(QWQ^T) \leq 1 \right\}$$

(for proof, see Lemma 14.3.7).

Separable polynomial perturbations. Here the structure of perturbations is as follows: ζ runs through the box $\mathcal{Z} = \{\zeta \in \mathbb{R}^L : \|\zeta\|_\infty \leq 1\}$, and the uncertain data entries are of the form

$$a = p_1^a(\zeta_1) + ... + p_L^a(\zeta_L),$$

where $p_\ell^a(s)$ are given algebraic polynomials of degrees not exceeding d; in other words, the basic functions can be split into L groups, the functions of ℓ-th group being $1 = \zeta_\ell^0, \zeta_\ell, \zeta_\ell^2, ..., \zeta_\ell^d$. Consequently, the function $\widehat{\zeta}[\zeta]$ is given by

$$\widehat{\zeta}[\zeta] = [[1; \zeta_1; \zeta_1^2; ...; \zeta_1^d]; ...; [1; \zeta_L; \zeta_L^2; ...; \zeta_L^d]].$$

Setting $P = \{\widehat{s} = [1; s; s^2; ...; s^d] : -1 \leq s \leq 1\}$, we conclude that $\widetilde{\mathcal{Z}} = \widehat{\zeta}[\mathcal{Z}]$ can be identified with the set $P^L = \underbrace{P \times ... \times P}_{L}$, so that $\widehat{\mathcal{Z}}$ is nothing but the set

$\underbrace{\mathcal{P}\times ... \times\mathcal{P}}_{L}$, where $\mathcal{P} = \mathrm{Conv}(P)$. It remains to note that the set $\mathcal{P}$ admits an explicit semidefinite representation, see Lemma 14.3.4.

1.5 EXERCISES

Exercise 1.1. Consider an uncertain LO problem with instances

$$\min_x \left\{c^T x : Ax \leq b\right\} \qquad [A : m\times n]$$

and with simple interval uncertainty:

$$\mathcal{U} = \{(c, A, b) : |c_j - c_j^{\mathrm{n}}| \leq \sigma_j, |A_{ij} - A_{ij}^{\mathrm{n}}| \leq \alpha_{ij}, |b_i - b_i^{\mathrm{n}}| \leq \beta_i \forall i, j\}$$

($^{\mathrm{n}}$ marks the nominal data). Reduce the RC of the problem to an LO problem with m constraints (not counting the sign constraints on the variables) and $2n$ nonnegative variables.

Exercise 1.2. Represent the RCs of every one of the uncertain linear constraints given below:

$$\begin{array}{ll} a^T x \leq b, [a; b] \in \mathcal{U} = \{[a; b] = [a^{\mathrm{n}}; b^{\mathrm{n}}] + P\zeta : \|\zeta\|_p \leq \rho\} & \\ \qquad\qquad\qquad\qquad\qquad [p \in [1, \infty]] & (a) \\ a^T x \leq b, [a; b] \in \mathcal{U} = \{[a; b] = [a^{\mathrm{n}}; b^{\mathrm{n}}] + P\zeta : \|\zeta\|_p \leq \rho, \zeta \geq 0\} & \\ \qquad\qquad\qquad\qquad\qquad [p \in [1, \infty]] & (b) \\ a^T x \leq b, [a; b] \in \mathcal{U} = \{[a; b] = [a^{\mathrm{n}}; b^{\mathrm{n}}] + P\zeta : \|\zeta\|_p \leq \rho\} & \\ \qquad\qquad\qquad\qquad\qquad [p \in (0, 1)] & (c) \end{array}$$

as explicit convex constraints.

Exercise 1.3. Represent in tractable form the RC of uncertain linear constraint

$$a^T x \leq b$$

with $\cap$-ellipsoidal uncertainty set

$$\mathcal{U} = \{[a, b] = [a^{\mathrm{n}}; b^{\mathrm{n}}] + P\zeta : \zeta^T Q_j \zeta \leq \rho^2, 1 \leq j \leq J\},$$

where $Q_j \succeq 0$ and $\sum_j Q_j \succ 0$.

1.6 NOTES AND REMARKS

NR 1.1. The paradigm of Robust Linear Optimization in the form considered here goes back to A.L. Soyster [109], 1973. To the best of our knowledge, in two subsequent decades there were only two publications on the subject [52, 106]. The activity in the area was revived circa 1997, independently and essentially simultaneously, in the frameworks of both Integer Programming (Kouvelis and Yu [70]) and Convex Programming (Ben-Tal and Nemirovski [3, 4], El Ghaoui et al. [49, 50]). Since 2000, the RO area is witnessing a burst of research activity in both theory and applications, with numerous researchers involved worldwide. The magnitude and diversity of the related contributions make it beyond our abilities to discuss them here. The reader can get some impression of this activity from [9, 16, 110, 89] and references therein.

NR 1.2. By itself, the RO methodology can be applied to every optimization problem where one can separate numerical data (that can be partly uncertain) from a problem's structure (that is known in advance and common for all instances of the uncertain problem). In particular, the methodology is fully applicable to uncertain *mixed integer* LO problems, where part of the decision variables are restricted to be integer. Note, however, that tractability issues, (which are our main focus in this book), in Uncertain LO with real variables and Uncertain Mixed-Integer LO need quite different treatment. While Theorem 1.3.4 is fully applicable to the mixed integer case and implies, in particular, that the RC of an uncertain mixed-integer LO problem $\mathcal{P}$ with a polyhedral uncertainty set is an explicit mixed-integer LO program with exactly the same integer variables as those of the instances of $\mathcal{P}$, the "tractability consequences" of this fact are completely different from those we made in the main body of this chapter. With no integer variables, the fact that the RC is an LO program straightforwardly implies tractability of the RC, while in the presence of integer variables no such conclusion can be made. Indeed, in the mixed integer case already the instances of the uncertain problem $\mathcal{P}$ typically are intractable, which, of course, implies intractability of the RC. In the case when the instances of $\mathcal{P}$ are tractable, the "fine structure" of the instances responsible for this rare phenomenon usually is destroyed when passing to the mixed-integer reformulation of the RC. There are some remarkable exceptions to this rule (see, e.g., [25]); however, in general the Uncertain Mixed-Integer LO is incomparably more complex computationally than the Uncertain LO with real variables. As it was already stated, our book is primarily focused on tractability issues of RO, and in order to get positive results in this direction, we restrict ourselves to uncertain problems with well-structured convex (and thus tractable) instances.

NR 1.3. Tractability of the RC of an uncertain LO problem with a tractable uncertainty set was established in the very first papers on convex RO. Theorem 1.3.4 and Corollary 1.3.5 are taken from [5].

Chapter Two

Robust Counterpart Approximations of Scalar Chance Constraints

2.1 HOW TO SPECIFY AN UNCERTAINTY SET

The question posed in the title of this section goes beyond general-type theoretical considerations — this is mainly a modeling issue that should be resolved on the basis of application-driven considerations. There is however a special case where this question makes sense and can, to some extent, be answered — this is the case where our goal is not to build an uncertainty model "from scratch," but rather to *translate* an already existing uncertainty model, namely, a stochastic one, to the language of "uncertain-but-bounded" perturbation sets and the associated robust counterparts. By exactly the same reasons as in the previous section, we can restrict our considerations to the case of a *single* uncertainty-affected linear inequality (1.3.4), (1.3.5).

Probabilistic vs. "uncertain-but-bounded" perturbations. When building the RC (1.3.6) of uncertain linear inequality (1.3.4), we worked with the so called "uncertain-but-bounded" data model (1.3.5) — one where all we know about the possible values of the data $[a; b]$ is their domain $\mathcal{U}$ defined in terms of a given affine parameterization of the data by perturbation vector ζ varying in a given perturbation set $\mathcal{Z}$. It should be stressed that we did not assume that the perturbations are of a stochastic nature and therefore used the only approach meaningful under the circumstances, namely, we looked for solutions that remain feasible whatever the data perturbation from $\mathcal{Z}$. This approach has its advantages:

i) More often than not there are no reasons to assign the perturbations a stochastic nature.

Indeed, stochasticity makes sense only when one repeats a certain action many times, or executes many similar actions in parallel; here it might be reasonable to think of frequencies of successes, etc. Probabilistic considerations become, methodologically, much more problematic when applied to a unique action, with no second attempt possible.

ii) Even when the unknown data can be thought of as stochastic, it might be difficult, especially in the large-scale case, to specify reliably data distribution. Indeed, the mere fact that the data are stochastic does not help unless we possess at least a partial knowledge of the underlying distribution.

> Of course, the uncertain-but-bounded models of uncertainty also require a priori knowledge, namely, to know what is the uncertainty set (a probabilistically oriented person could think about this set as the *support* of data distribution, that is, the smallest closed set in the space of the data such that the probability for the data to take a value outside of this set is zero). Note, however, that it is much easier to point out the support of the relevant distribution than the distribution itself.

With the uncertain-but-bounded model of uncertainty, we can make clear predictions like "with such and such behavior, we definitely will survive, provided that the unknown parameters will differ from their nominal values by no more than 15%, although we may die when the variation will be as large as 15.1%." In case we do believe that 15.1% variations are also worthy to worry about, we have an option to increase the perturbation set to take care of 30% perturbations in the data. With luck, we will be able to find a robust feasible solution for the increased perturbation set. This is a typical engineering approach — after the required thickness of a bar supporting certain load is found, a civil engineer will increase it by factor like 1.2 or 1.5 "to be on the safe side" — to account for model inaccuracies, material imperfections, etc. With a stochastic uncertainty model, this "being on the safe side" is impossible — increasing the probability of certain events, one must decrease simultaneously the probability of certain other events, since the "total probability budget" is once and for ever fixed. While all these arguments demonstrate that there are situations in reality when the uncertain-but-bounded model of data perturbations possesses significant methodological advantages over the stochastic models of uncertainty, there are, of course, applications (like communications, weather forecasts, mass production, and, to some extent, finance) where one can rely on probabilistic models of uncertainty. Whenever this is the case, the much less informative uncertain-but-bounded model and associated worst-case-oriented decisions can be too conservative and thus impractical. The bottom line is that *while the stochastic models of data uncertainty are by far not the only meaningful ones, they definitely deserve attention.* Our goal in this chapter is to *develop techniques that are capable to utilize, to some extent, knowledge of the stochastic nature of data perturbations when building uncertainty-immunized solutions.* This goal will be achieved via a specific "translation" of stochastic models of uncertain data to the language of uncertain-but-bounded perturbations and the associated robust counterparts. Before developing the approach in full detail, we will explain why we choose such an implicit way to treat stochastic uncertainty models instead of treating them directly.

2.2 CHANCE CONSTRAINTS AND THEIR SAFE TRACTABLE APPROXIMATIONS

The most direct way to treat stochastic data uncertainty in the context of uncertain Linear Optimization is offered by an old concept (going back to 50s [40]) of *chance* *constraints.* Consider an uncertain linear inequality

$$a^T x \leq b, \ [a; b] = [a^0; b^0] + \sum_{\ell=1}^{L} \zeta_\ell [a^\ell; b^\ell] \tag{2.2.1}$$

(cf. (1.3.4), (1.3.5)) and assume that the perturbation vector ζ is random with, say, completely known probability distribution P. Ideally, we would like to work with candidate solutions x that make the constraint valid with probability 1. This "ideal goal," however, means coming back to the uncertain-but-bounded model of perturbations; indeed, it is easily seen that a given x satisfies (2.2.1) for almost all realizations of ζ if and only if x is robust feasible w.r.t. the perturbation set that is the closed convex hull of the support of P. The only meaningful way to utilize the stochasticity of perturbations is to require a candidate solution x to satisfy the constraint for "nearly all" realizations of ζ, specifically, to satisfy the constraint with probability at least $1 - \epsilon$, where $\epsilon \in (0, 1)$ is a prespecified small tolerance. This approach associates with the randomly perturbed constraint (2.2.1) the *chance constraint*

$$p(x) \equiv \text{Prob}_{\zeta \sim P} \left\{ \zeta : [a^0]^T x + \sum_{\ell=1}^{L} \zeta_\ell [a^\ell]^T x > b^0 + \sum_{\ell=1}^{L} \zeta_\ell b^\ell \right\} \leq \epsilon, \tag{2.2.2}$$

where $\text{Prob}_{\zeta \sim P}$ is the probability associated with the distribution P. Note that (2.2.2) is a usual certain constraint. Replacing all uncertainty-affected constraints in an uncertain LO problem with their chance constrained versions and minimizing the objective function, (which we, w.l.o.g., may assume to be certain) under these constraints, we end up with the *chance constrained* version of $(\text{LO}_\mathcal{U})$, which is a deterministic optimization problem.

While the outlined approach seems to be quite natural, it suffers from a severe drawback — *typically, it results in a severely computationally intractable problem.* The reason is twofold:

i) Usually, it is difficult to evaluate with high accuracy the probability in the left hand side of (2.2.2), even in the case when P is simple.

For example, it is known [68] that computing the left hand side in (2.2.2) is NP-hard already when ζ_ℓ are independent and uniformly distributed in $[-1, 1]$. This means that unless P=NP, there is no algorithm that, given on input a rational x, rational data $\{[a^\ell; b^\ell]\}_{\ell=0}^{L}$ and rational $\delta \in (0, 1)$, allows to evaluate $p(x)$ within accuracy δ in time polynomial in the bit size of the input. Unless ζ takes values in a finite set of moderate cardinality, the only known general method to evaluate $p(x)$ is based on Monte-Carlo simulations; this method, however, requires samples with cardinality of order of $1/\delta$, where δ is the required accuracy of evaluation. Since the meaningful values of this accuracy are $\leq \epsilon$, we conclude that in reality the Monte-Carlo approach can hardly be used when ϵ is like 0.0001 or less.

ii) More often than not the feasible set of (2.2.2) is non-convex, which makes optimization under chance constraints a highly problematic task.

> Note that while the first difficulty becomes an actual obstacle only when ϵ is small enough, the second difficulty makes chance constrained optimization highly problematic for "large" ϵ as well.

Essentially, the only known case when none of the outlined difficulties occur is the case where ζ is a Gaussian random vector and $\epsilon < 1/2$.

Due to the severe computational difficulties associated with chance constraints, a natural course of action is to replace a chance constraint with its *computationally tractable safe approximation.* The latter notion is defined as follows:

Definition 2.2.1. Let $\{[a^\ell; b^\ell]\}_{\ell=0}^L$, P, ϵ be the data of chance constraint (2.2.2), and let $\mathcal{S}$ be a system of convex constraints on x and additional variables v. We say that $\mathcal{S}$ is a safe convex approximation of chance constraint (2.2.2), if the x component of every feasible solution (x, v) of $\mathcal{S}$ is feasible for the chance constraint.

A safe convex approximation $\mathcal{S}$ of (2.2.2) is called computationally tractable, if the convex constraints forming $\mathcal{S}$ are efficiently computable.

It is clear that by replacing the chance constraints in a given chance constrained optimization problem with their safe convex approximations, we end up with a convex optimization problem in x and additional variables that is a "safe approximation" of the chance constrained problem: the x component of every feasible solution to the approximation is feasible for the chance constrained problem. If the safe convex approximation in question is tractable, then the above approximating program is a convex program with efficiently computable constraints and as such it can be processed efficiently.

In the sequel, when speaking about safe convex approximations, we omit for the sake of brevity the adjective "convex," which should always be added "by default."

2.2.1 Ambiguous Chance Constraints

Chance constraint (2.2.2) is associated with randomly perturbed constraint (2.2.1) *and a given distribution P of random perturbations*, and it is reasonable to use this constraint when we do know this distribution. In reality we usually have only *partial* information on P, that is, we know only that P belongs to a given family $\mathcal{P}$ of distributions. When this is the case, it makes sense to pass from (2.2.2) to the *ambiguous* chance constraint

$$\forall(P \in \mathcal{P}) : \mathrm{Prob}_{\zeta\sim P}\left\{\zeta : [a^0]^T x + \sum_{\ell=1}^{L} \zeta_\ell [a^\ell]^T x > b^0 + \sum_{\ell=1}^{L} \zeta_\ell b^\ell\right\} \leq \epsilon. \qquad (2.2.3)$$

Of course, the definition of a safe tractable approximation of chance constraint extends straightforwardly to the case of ambiguous chance constraint. In the sequel, we usually skip the adjective "ambiguous"; what exactly is meant depends on whether we are speaking about a partially or a fully known distribution P.

Next we present a simple scheme for the safe approximation of chance constraints.

2.3 SAFE TRACTABLE APPROXIMATIONS OF SCALAR CHANCE CONSTRAINTS: BASIC EXAMPLES

Consider the case of chance constraint (2.2.3) where all we know about the random variables ζ_ℓ is that

$$\mathbf{E}\{\zeta_\ell\} = 0 \ \& \ |\zeta_\ell| \leq 1, \ \ell = 1, ..., L \ \& \ \{\zeta_\ell\}_{\ell=1}^L \text{ are independent} \tag{2.3.1}$$

(that is, $\mathcal{P}$ is comprised of all distributions satisfying (2.3.1)). Note that a more general case of independent random variables ζ_ℓ taking values in given finite segments centered at the expectations of ζ_ℓ by "scalings" $\zeta_\ell \mapsto \xi_\ell = \alpha_\ell \zeta_\ell + \beta_\ell$ with deterministic α_ℓ, β_ℓ can be reduced to (2.3.1) (cf. Remark 1.2.2).

Observe that the body of chance constraint (2.2.2) can be rewritten as

$$\eta \equiv \sum_{\ell=1}^{L} [[a^\ell]^T x - b^\ell] \zeta_\ell \leq b^0 - [a^0]^T x. \tag{2.3.2}$$

In the case of (2.3.1), for x fixed, η is a random variable with zero mean and standard deviation

$$\mathrm{StD}[\eta] = \sqrt{\sum_{\ell=1}^{L} ([a^\ell]^T x - b^\ell)^2 \mathbf{E}\{\zeta_\ell^2\}} \leq \sqrt{\sum_{\ell=1}^{L} ([a^\ell]^T x - b^\ell)^2}.$$

The chance constraint requires for (2.3.2) to be satisfied with probability $\geq 1 - \epsilon$. An engineer would respond to this requirement arguing that a random variable is "never" greater than its mean plus 3 times the standard deviation, so that η is "never" greater than the quantity $3\sqrt{\sum_{\ell=1}^{L} ([a^\ell]^T x - b^\ell)^2}$. We need not be as specific as an engineer and say that η is "nearly never" greater than the quantity $\Omega\sqrt{\sum_{\ell=1}^{L} ([a^\ell]^T x - b^\ell)^2}$, where Ω is a "safety parameter" of order of 1; the larger Ω, the less the chances for η to be larger than the outlined quantity. We thus arrive at a parametric "safe" version

$$\Omega\sqrt{\sum_{\ell=1}^{L} ([a^\ell]^T x - b^\ell)^2} \leq b^0 - [a^0]^T x \tag{2.3.3}$$

of the randomly perturbed constraint (2.3.2). It seems that with properly defined Ω, every feasible solution to this constraint satisfies, with probability at least $1 - \epsilon$, the inequality in (2.3.2). This indeed is the case; a simple analysis, which we will carry later on, demonstrates that our "engineering reasoning" can be justified.

Specifically, the following is true (for proof see Remark 2.4.10 and Proposition 2.4.2):

Proposition 2.3.1. Let z_ℓ, $\ell = 1, ..., L$, be deterministic coefficients and ζ_ℓ, $\ell = 1, ..., L$, be independent random variables with zero mean taking values in $[-1, 1]$. Then for every $\Omega \geq 0$ it holds that

$$\text{Prob}\left\{\zeta : \sum_{\ell=1}^{L} z_\ell \zeta_\ell > \Omega \sqrt{\sum_{\ell=1}^{L} z_\ell^2}\right\} \leq \exp\{-\Omega^2/2\}. \tag{2.3.4}$$

As an immediate conclusion, we get

$$(2.3.1) \Rightarrow \text{Prob}\left\{\eta > \Omega \sqrt{\sum_{\ell=1}^{L}([a^\ell]^T x - b^\ell)^2}\right\} \leq \exp\{-\Omega^2/2\}\ \forall \Omega \geq 0, \tag{2.3.5}$$

and we have arrived at the result as follows.

Corollary 2.3.2. In the case of (2.3.1), the conic quadratic constraint (2.3.3) is a computationally tractable safe approximation of the chance constraint

$$\text{Prob}\left\{[a^0]^T x + \sum_{\ell=1}^{L} \zeta_\ell [a^\ell]^T x > b^0 + \sum_{\ell=1}^{L} \zeta_\ell b^\ell\right\} \leq \exp\{-\Omega^2/2\}. \tag{2.3.6}$$

In particular, with $\Omega \geq \sqrt{2\ln(1/\epsilon)}$, the constraint (2.3.3) is a tractable safe approximation of the chance constraint (2.2.2).

Now let us make the following important observation:

> In view of Example 1.3.3, *inequality* (2.3.3) *is nothing but the RC of the uncertain linear inequality* (2.2.1), (1.3.5), *with the perturbation set* $\mathcal{Z}$ *in* (1.3.5) *specified as the ball*
>
> $$\text{Ball}_\Omega = \{\zeta : \|\zeta\|_2 \leq \Omega\}. \tag{2.3.7}$$

This observation is worthy of in-depth discussion.

A. By itself, the assumption that ζ_ℓ vary in $[-1, 1]$, (which is a part of the assumptions in (2.3.1)), suggests to consider, as the perturbation set $\mathcal{Z}$ in (1.3.5), the box

$$\text{Box}_1 = \{\zeta : -1 \leq \zeta_\ell \leq 1,\ \ell = 1, ..., L\}.$$

For this $\mathcal{Z}$, the associated RC of the uncertain linear inequality (2.2.1), (1.3.5) is

$$\sum_{\ell=1}^{L} |[a^\ell]^T x - b^\ell]| \leq b^0 - [a^0]^T x \tag{2.3.8}$$

(see Example 1.3.2). In the case of (2.3.1), this "box RC" guarantees "100% immunization against perturbations," meaning that every feasible solution to the box RC is feasible for the randomly perturbed inequality in question with probability 1. With the same stochastic model of uncertainty (2.3.1), the "ball RC," that is, the conic constraint (2.3.3), guarantees less, namely, "$(1-\exp\{-\Omega^2/2\})\cdot 100\%$-immunization." Note that with quite a moderate Ω, the "unreliability" $\exp\{-\Omega^2/2\}$ is negligible: it is less than 10^{-6} for $\Omega = 5.26$ and less than 10^{-12} for $\Omega = 7.44$. For all practical purposes, probability like 10^{-12} is the same as zero probability, so that there are all reasons to claim that the ball RC with $\Omega = 7.44$ is as "practically reliable" as the box RC.[1] Given that the "immunization power" of both RCs is essentially the same, it is very instructive to compare the "sizes" of the underlying perturbation sets Box_1 and Ball_Ω. This comparison leads to a great surprise: *when the dimension L of the perturbation set is not too small, the ball* Ball_Ω *with* Ω *"of order of one," say,* $\Omega = 7.44$, *is incomparably smaller that the unit box* Box_1 *with respect to all natural size measures such as diameter, volume, etc.* For example,

- the Euclidean diameters of Ball_Ω and Box_1 are respectively, 2Ω and $2\sqrt{L}$; with $\Omega = 7.44$, the second diameter is larger than the first starting with $L = 56$, and the ratio of the second diameter to the first one blows up to ∞ as L grows;

- the ratio of volumes of the ball and the box is

$$\frac{\mathrm{Vol}(\mathrm{Ball}_\Omega)}{\mathrm{Vol}(\mathrm{Box}_1)} = \frac{(\Omega\sqrt{\pi})^L}{2^L\Gamma(L/2+1)} \le \left(\frac{\Omega\sqrt{e\pi/2}}{\sqrt{L}}\right)^L,$$

Γ being the Euler Gamma function. For $\Omega = 7.44$, this ratio is < 1 starting with $L = 237$ and goes to 0 super-exponentially fast at $L \to \infty$.

B. As a counter-argument to what was said in **A**, one can argue that for small L the uncertainty set $\mathrm{Ball}_{7.44}$ is essentially larger than the uncertainty set Box_1. Well, here is a "rectification," interesting by its own right, of the ball RC which nullifies this counter-argument. Consider the case when the perturbation set $\mathcal{Z}$ is the intersection of the unit box and the ball of radius Ω centered at the origin:

$$\mathcal{Z} = \{\zeta \in \mathbb{R}^L : \|\zeta\|_\infty \le 1, \|\zeta\|_2 \le \Omega\} = \mathrm{Box}_1 \cap \mathrm{Ball}_\Omega. \tag{2.3.9}$$

Proposition 2.3.3. The RC of the uncertain linear constraint (2.2.1) with the uncertainty set (2.3.9) is equivalent to the system of conic quadratic constraints

$$\begin{array}{ll} (a) & z_\ell + w_\ell = b^\ell - [a^\ell]^T x,\ \ell = 1, ..., L; \\ (b) & \sum\limits_\ell |z_\ell| + \Omega\sqrt{\sum\limits_\ell w_\ell^2} \le b^0 - [a^0]^T x. \end{array} \tag{2.3.10}$$

In the case of (2.3.1), the x component of every feasible solution to this system satisfies the randomly perturbed inequality (2.2.1) with probability at least $1 - \exp\{-\Omega^2/2\}$.

Proof. The fact that (2.3.10) represents the RC of (2.2.1), the perturbation set being (2.3.9), is readily given by Example 1.3.7 where one should set $\sigma_\ell \equiv 1$. Now let us prove that if (2.3.1) takes place and x, z, w is feasible for (2.3.10), then x is feasible for (2.2.1) with probability at least $1 - \exp\{-\Omega^2/2\}$. Indeed, when

[1] This conclusion tacitly assumes that the underlying stochastic uncertainty model is accurate enough to be trusted even when speaking about probabilities as small as 1.e-12; concerns of this type seem to be the inevitable price for using stochastic models of uncertainty.

$\|\zeta\|_\infty \leq 1$, we have

$$\begin{array}{lll} & \sum_{\ell=1}^{L}[[a^\ell]^T x - b^\ell]\zeta_\ell > b^0 - [a^0]^T x & \\ \Rightarrow & -\sum_{\ell=1}^{L} z_\ell\zeta_\ell - \sum_{\ell=1}^{L} w_\ell\zeta_\ell > b^0 - [a^0]^T x & \text{[by (2.3.10.}a\text{)]} \\ \Rightarrow & \sum_{\ell=1}^{L}|z_\ell| - \sum_{\ell=1}^{L} w_\ell\zeta_\ell > b^0 - [a^0]^T x & \text{[since } \|\zeta\|_\infty \leq 1] \\ \Rightarrow & -\sum_{\ell=1}^{L} w_\ell\zeta_\ell > \Omega\sqrt{\sum_{\ell=1}^{L} w_\ell^2} & \text{[by (2.3.10.}b\text{)]} \end{array}$$

Therefore for every distribution P compatible with (2.3.1) we have

$$\mathrm{Prob}_{\zeta\sim P}\left\{x \text{ is ineasible for (2.2.1)}\right\} \leq \mathrm{Prob}_{\zeta\sim P}\left\{-\sum_{\ell=1}^{L} w_\ell\zeta_\ell > \Omega\sqrt{\sum_{\ell=1}^{L} w_\ell^2}\right\}$$
$$\leq \exp\{-\Omega^2/2\},$$

where the last inequality is due to Proposition 2.3.1. □

Note that perturbation set (2.3.9) is *never* greater than the perturbation set Box_1 and, as was explained in **A**, for every fixed Ω is incomparably smaller than the latter set when the dimension L of the perturbation vector ζ is large. Nevertheless, Proposition 2.3.3 says that when the perturbation vector is random and obeys (2.3.1), the "immunization power" of the RC associated with the small perturbation set (2.3.9), where $\Omega = 7.44$, is essentially as strong as that of the 100% reliable box RC (2.3.8). This phenomenon becomes even more striking when we consider the following special case of (2.2.1): ζ_ℓ are independent and each of them takes values ± 1 with probabilities $1/2$. In this case, when $L > \Omega^2$, the perturbation set (2.3.9) does not contain even a *single* realization of the random perturbation vector! Thus, the "immunization power" of the RC (2.3.10) cannot be explained by the fact that the underlying perturbation set contains "nearly all" realizations of the random perturbation vector.

C. Our considerations justify the use of "strange" perturbation sets like ellipsoids and intersections of ellipsoids and parallelotopes: while it may seem difficult to imagine a natural perturbation mechanism that produces perturbations from such sets, our analysis demonstrates that these sets do emerge naturally when "immunizing" solutions against random perturbations of the type described in (2.3.1). The same is true for the "budgeted" perturbation set considered in Example 1.3.8:

Proposition 2.3.4. Consider the RC of uncertain linear constraint (2.2.1) in the case of budgeted uncertainty:

$$\mathcal{Z} = \{\zeta \in \mathbb{R}^L : -1 \leq \zeta_\ell \leq 1, \ell = 1, ..., L, \sum_{\ell=1}^{L} |\zeta_\ell| \leq \gamma\}. \tag{2.3.11}$$