Lecture 5

Adjustable Robust Multistage Optimization

In this lecture we intend to investigate robust *multi-stage* linear and conic optimization.

5.1 Adjustable Robust Optimization: Motivation

Consider a general-type uncertain optimization problem — a collection

$$\mathcal{P} = \left\{ \min_x \{ f(x,\zeta) : F(x,\zeta) \in \mathbf{K} \} : \zeta \in \mathcal{Z} \right\} \tag{5.1.1}$$

of *instances* — optimization problems of the form

$$\min_x \{ f(x,\zeta) : F(x,\zeta) \in \mathbf{K} \},$$

where $x \in \mathbb{R}^n$ is the decision vector, $\zeta \in \mathbb{R}^L$ represents the uncertain data or data perturbation, the real-valued function $f(x,\zeta)$ is the objective, and the vector-valued function $F(x,\zeta)$ taking values in $\mathbb{R}^m$ along with a set $\mathbf{K} \subset \mathbb{R}^m$ specify the constraints; finally, $\mathcal{Z} \subset \mathbb{R}^L$ is the uncertainty set where the uncertain data is restricted to reside.

> Format (5.1.1) covers all uncertain optimization problems considered so far; moreover, in these latter problems the objective f and the right hand side F of the constraints always were *bi-affine* in x, ζ, (that is, affine in x when ζ is fixed, and affine in ζ, x being fixed), and $\mathbf{K}$ was a "simple" convex cone (a direct product of nonnegative rays/Lorentz cones/Semidefinite cones, depending on whether we were speaking about uncertain Linear, Conic Quadratic or Semidefinite Optimization). We shall come back to this "well-structured" case later; for our immediate purposes the specific conic structure of instances plays no role, and we can focus on "general" uncertain problems in the form of (5.1.1).

The Robust Counterpart of uncertain problem (5.1.1) is defined as the semi-infinite optimization problem

$$\min_{x,t} \{ t : \forall \zeta \in \mathcal{Z} : f(x,\zeta) \le t, F(x,\zeta) \in \mathbf{K} \}; \tag{5.1.2}$$

this is exactly what was so far called the RC of an uncertain problem.

Recall that our interpretation of the RC (5.1.2) as the natural source of robust/robust optimal solutions to the uncertain problem (5.1.1) is not self-evident, and its "informal justification" relies upon the specific assumptions A.1–3 on our "decision environment," see page 8. We have already relaxed somehow the last of these assumptions, thus arriving at the notion of Globalized Robust Counterpart, lecture 4. What is on our agenda now is to revise the first assumption, which reads

> A.1. All decision variables in (5.1.1) represent "here and now" decisions; they should get specific numerical values as a result of solving the problem *before* the actual data "reveals itself" and as such should be independent of the actual values of the data.

We have considered numerous examples of situations where this assumption is valid. At the same time, there are situations when it is too restrictive, since "in reality" some of the decision variables can adjust themselves, to some extent, to the actual values of the data. One can point out at least two sources of such adjustability: presence of *analysis variables* and *wait-and-see decisions*.

Analysis variables. Not always all decision variables x_j in (5.1.1) represent actual decisions; in many cases, some of x_j are slack, or analysis, variables introduced in order to convert the instances into a desired form, e.g., the one of Linear Optimization programs. It is very natural to allow for the analysis variables to depend on the true values of the data — why not?

Example 5.1 [cf. Example 1.3] Consider an "ℓ_1 constraint"

$$\sum_{k=1}^{K} |a_k^T x - b_k| \le \tau; \tag{5.1.3}$$

you may think, e.g., about the Antenna Design problem (Example 1.1) where the "fit" between the actual diagram of the would-be antenna array and the target diagram is quantified by the $\|\cdot\|_1$ distance. Assuming that the data and x are real, (5.1.3) can be represented equivalently by the system of linear inequalities

$$-y_k \le a_k^T x - b_k \le y_k, \ \sum_k y_k \le \tau$$

in variables x, y, τ. Now, when the data a_k, b_k are uncertain and the components of x do represent "here and now" decisions and should be independent of the actual values of the data, there is absolutely no reason to impose the latter requirement on the slack variables y_k as well: they do not represent decisions at all and just certify the fact that the actual decisions x, τ meet the requirement (5.1.3). While we can, of course, impose this requirement "by force," this perhaps will lead to a too conservative model. It seems to be completely natural to allow for the certificates y_k to depend on actual values of the data — it may well happen that then we shall be able to certify robust feasibility for (5.1.3) for a larger set of pairs (x, τ).

Wait-and-see decisions. This source of adjustability comes from the fact that some of the variables x_j represent decisions that are not "here and now" decisions, i.e., those that should be made before the true data "reveals itself." In multi-stage decision making processes, some x_j can represent "wait and see" decisions, which could be made after the controlled system "starts to live," at time instants when part (or all) of the true data is revealed. It is fully legitimate to allow for these decisions to depend on the part of the data that indeed "reveals itself" before the decision should be made.

Example 5.2 Consider a multi-stage inventory system affected by uncertain demand. The most interesting of the associated decisions — the *replenishment orders* — are made one at a time, and the replenishment order of "day" t is made when we already know the actual demands in the preceding days. It is completely natural to allow for the orders of day t to depend on the preceding demands.

5.2 Adjustable Robust Counterpart

A natural way to model adjustability of variables is as follows: for every $j \leq n$, we allow for x_j to depend on a prescribed "portion" $P_j\zeta$ of the true data ζ:

$$x_j = X_j(P_j\zeta), \tag{5.2.1}$$

where $P_1, ..., P_n$ are given in advance matrices specifying the "information base" of the decisions x_j, and $X_j(\cdot)$ are *decision rules* to be chosen; these rules can in principle be arbitrary functions on the corresponding vector spaces. For a given j, specifying P_j as the zero matrix, we force x_j to be completely independent of ζ, that is, to be a "here and now" decision; specifying P_j as the unit matrix, we allow for x_j to depend on the entire data (this is how we would like to describe the analysis variables). And the "in-between" situations, choosing P_j with $1 \leq \text{Rank}(P_j) < L$ enables one to model the situation where x_j is allowed to depend on a "proper portion" of the true data.

We can now replace in the usual RC (5.1.2) of the uncertain problem (5.1.1) the independent of ζ decision variables x_j with functions $X_j(P_j\zeta)$, thus arriving at the problem

$$\begin{array}{c} \min\limits_{t, \{X_j(\cdot)\}_{j=1}^n} \left\{ t : \forall \zeta \in \mathcal{Z} : f(X(\zeta), \zeta) \leq t, F(X(\zeta), \zeta) \in \mathbf{K} \right\}, \\ X(\zeta) = [X_1(P_1\zeta); ...; X_n(P_n\zeta)]. \end{array} \tag{5.2.2}$$

The resulting optimization problem is called the *Adjustable Robust Counterpart* (ARC) of the uncertain problem (5.1.1), and the (collections of) decision rules $X(\zeta)$, which along with certain t are feasible for the ARC, are called *robust feasible decision rules.* The ARC is then the problem of specifying a collection of decision rules with prescribed information base that is feasible for as small t as possible. The *robust optimal decision rules* now replace the *constant* (non-adjustable, data-independent) robust optimal decisions that are yielded by the usual Robust Counterpart (5.1.2) of our uncertain problem. Note that the ARC is an extension of the RC; the latter is a "trivial" particular case of the former corresponding to the case of trivial information base in which all matrices P_j are zero.

5.2.1 Examples

We are about to present two instructive examples of uncertain optimization problems with adjustable variables.

Information base induced by time precedences. In many cases, decisions are made subsequently in time; whenever this is the case, a natural information base of the decision to be made at instant t ($t = 1, ..., N$) is the part of the true data that becomes known at time t. As an instructive example, consider a simple Multi-Period Inventory model mentioned in Example 5.2:

Example 5.2 continued. Consider an inventory system where d products share common warehouse capacity, the time horizon is comprised of N periods, and the goal is to minimize the total inventory management cost. Allowing for backlogged demand, the simplest model of such an inventory looks as follows:

$$\begin{array}{lll}
\text{minimize} & C & \text{[inventory management cost]} \\
\text{s.t.} & & \\
(a) & C \geq \sum_{t=1}^{N} \left[c_{\mathrm{h},t}^T y_t + c_{\mathrm{b},t}^T z_t + c_{\mathrm{o},t}^T w_t \right] & \text{[cost description]} \\
(b) & x_t = x_{t-1} + w_t - \zeta_t,\ 1 \leq t \leq N & \text{[state equations]} \\
(c) & y_t \geq 0, y_t \geq x_t,\ 1 \leq t \leq N & \\
(d) & z_t \geq 0, z_t \geq -x_t,\ 1 \leq t \leq N & \\
(e) & \underline{w}_t \leq w_t \leq \overline{w}_t,\ 1 \leq t \leq N & \\
(f) & q^T y_t \leq r &
\end{array} \tag{5.2.3}$$

The variables in this problem are:

- $C \in \mathbb{R}$ — (upper bound on) the total inventory management cost;
- $x_t \in \mathbb{R}^d$, $t = 1, ..., N$ — states. i-th coordinate x_t^i of vector x_t is the amount of product of type i that is present in the inventory at the time instant t (end of time interval # t). This amount can be nonnegative, meaning that the inventory at this time has x_t^i units of free product # i; it may be also negative, meaning that the inventory at the moment in question owes the customers $|x_t^i|$ units of the product i ("backlogged demand"). The initial state x_0 of the inventory is part of the data, and not part of the decision vector;
- $y_t \in \mathbb{R}^d$ are upper bounds on the positive parts of the states x_t, that is, (upper bounds on) the "physical" amounts of products stored in the inventory at time t, and the quantity $c_{\mathrm{h},t}^T y_t$ is the (upper bound on the) holding cost in the period t; here $c_{\mathrm{h},t} \in \mathbb{R}_+^d$ is a given vector of the holding costs per unit of the product. Similarly, the quantity $q^T y_t$ is (an upper bound on) the warehouse capacity used by the products that are "physically present" in the inventory at time t, $q \in \mathbb{R}_+^d$ being a given vector of the warehouse capacities per units of the products;
- $z_t \in \mathbb{R}^d$ are (upper bounds on) the backlogged demands at time t, and the quantities $c_{\mathrm{b},t}^T z_t$ are (upper bounds on) the penalties for these backlogged demands. Here $c_{\mathrm{b,t}} \in \mathbb{R}_+^d$ are given vectors of the penalties per units of the backlogged demands;
- $w_t \in \mathbb{R}^d$ is the vector of replenishment orders executed in period t, and the quantities $c_{\mathrm{o},t}^T w_t$ are the costs of executing these orders. Here $c_{\mathrm{o},t} \in \mathbb{R}_+^d$ are given vectors of per unit ordering costs.

With these explanations, the constraints become self-evident:

- (a) is the "cost description": it says that the total inventory management cost is comprised of total holding and ordering costs and of the total penalty for the backlogged demand;
- (b) are state equations: "what will be in the inventory at the end of period t (x_t) is what was there at the end of preceding period (x_{t-1}) plus the replenishment orders of the period (w_t) minus the demand of the period (ζ_t);
- (c), (d) are self-evident;
- (e) represents the upper and lower bounds on replenishment orders, and (f) expresses the requirement that (an upper bound on) the total warehouse capacity $q^T y_t$ utilized by products that are "physically present" in the inventory at time t should not be greater than the warehouse capacity r.

In our simple example, we assume that out of model's parameters

$$x_0, \{c_{\mathrm{h},t}, c_{\mathrm{b},t}, c_{\mathrm{o,t}}, \underline{w}_t, \overline{w}_t\}_{t=1}^N, q, r, \{\zeta_t\}_{t=1}^N$$

the only uncertain element is the *demand trajectory* $\zeta = [\zeta_1; ...; \zeta_N] \in \mathbb{R}^{dN}$, and that this trajectory is known to belong to a given uncertainty set $\mathcal{Z}$. The resulting uncertain Linear Optimization problem is comprised of instances (5.2.3) parameterized by the uncertain data — demand trajectory ζ — running through a given set $\mathcal{Z}$.

As far as the adjustability is concerned, all variables in our problem, except for the replenishment orders w_t, are analysis variables. As for the orders, the simplest assumption is that w_t should get numerical value at time t, and that at this time we already know the past demands $\zeta^{t-1} = [\zeta_1; ...; \zeta_{t-1}]$. Thus, the information base for w_t is $\zeta^{t-1} = P_t\zeta$ (with the convention that $\zeta^s = 0$ when $s < 0$). For the remaining analysis variables the information base is the entire demand trajectory ζ. Note that we can easily adjust this model to the case when there are lags in demand acquisition, so that w_t should depend on a prescribed initial segment $\zeta^{\tau(t)-1}$, $\tau(t) \leq t$, of ζ^{t-1} rather than on the entire ζ^{t-1}. We can equally easily account for the possibility, if any, to observe the demand "on line," by allowing w_t to depend on ζ^t rather than on ζ^{t-1}. Note that in all these cases the information base of the decisions is readily given by the natural time precedences between the "actual decisions" augmented by a specific demand acquisition protocol.

Example 5.3 Project management. Figure 5.1 is a simple *PERT diagram* — a graph representing a Project Management problem. This is an acyclic directed graph with nodes corresponding to *events*, and arcs corresponding to *activities*. Among the nodes there is a *start node* S with no incoming arcs and an *end node* F with no outgoing arcs, interpreted as "start of the project" and "completion of the project," respectively. The remaining nodes correspond to the events "a specific stage of the project is completed, and one can pass to another stage". For example, the diagram could represent creating a factory, with A, B, C being, respectively, the events "equipment to be installed is acquired and delivered," "facility #1 is built and equipped," "facility # 2 is built and equipped." The activities are jobs comprising the project. In our example, these jobs could be as follows:

a: acquiring and delivering the equipment for facilities ## 1,2
b: building facility # 1
c: building facility # 2
d: installing equipment in facility # 1
e: installing equipment in facility # 2
f: training personnel and preparing production at facility # 1
g: training personnel and preparing production at facility # 2

The topology of a PERT diagram represents *logical precedences* between the activities and events: a particular activity, say g, can start only after the event C occurs, and the latter event happens when both activities c and e are completed.

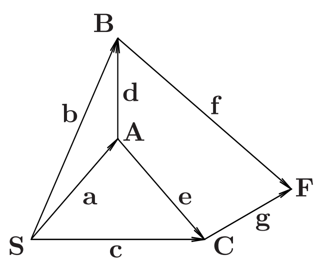


Figure 5.1: A PERT diagram.

In PERT models it is assumed that activities γ have nonnegative durations τ_γ (perhaps depending on control parameters), and are executed without interruptions, with possible idle periods between the moment when the start of an activity is allowed by the logical precedences and the moment when it is actually started. With these assumptions, one can write down a system of constraints on the time instants t_ν when events ν can take place. Denoting by $\Gamma = \{\gamma = (\mu_\gamma, \nu_\gamma)\}$ the set of arcs in a PERT diagram (μ_γ is the start- and ν_γ is the end-node of an arc γ), this system reads

$$t_{\mu_\gamma} - t_{\nu_\gamma} \geq \tau_\gamma \ \forall \gamma \in \Gamma. \tag{5.2.4}$$

"Normalizing" this system by the requirement

$$t_S = 0,$$

the values of t_F, which can be obtained from feasible solutions to the system, are achievable durations of the entire project. In a typical Project Management problem, one imposes an upper bound on t_F and minimizes, under this restriction, coupled with the system of constraints (5.2.4), some objective function.

As an example, consider the situation where the "normal" durations τ_γ of activities can be reduced at certain price ("in reality" this can correspond to investing into an activity extra manpower, machines, etc.). The corresponding model becomes

$$\tau_\gamma = \zeta_\gamma - x_\gamma, \ c_\gamma = f_\gamma(x_\gamma),$$

where ζ_γ is the "normal duration" of the activity, x_γ ("crush") is a nonnegative decision variable, and $c_\gamma = f_\gamma(x_\gamma)$ is the cost of the crush; here $f_\gamma(\cdot)$ is a given function. The associated optimization model might be, e.g., the problem of minimizing the total cost of the crushes under a given upper bound T on project's duration:

$$\min_{\substack{x=\{x_\gamma:\gamma\in\Gamma\}\\ \{t_\nu\}}} \left\{ \sum_\gamma f_\gamma(x_\gamma) : \begin{array}{l} t_{\mu_\gamma} - t_{\nu_\gamma} \geq \zeta_\gamma - x_\gamma \\ 0 \leq x_\gamma \leq \overline{x}_\gamma \end{array} \right\} \forall \gamma \in \Gamma, t_S = 0, t_F \leq T \Bigg\}, \tag{5.2.5}$$

where $\overline{x}_\gamma$ are given upper bounds on crushes. Note that when $f_\gamma(\cdot)$ are convex functions, (5.2.5) is an explicit convex problem, and when, in addition to convexity, $f_\gamma(\cdot)$ are piecewise linear, (which is usually the case in reality and which we assume from now on), (5.2.5) can be straightforwardly converted to a Linear Optimization program.

Usually part of the data of a PERT problem are uncertain. Consider the simplest case when the only uncertain elements of the data in (5.2.5) are the normal durations ζ_γ of the activities (their uncertainty may come from varying weather conditions, inaccuracies in estimating the forthcoming effort, etc.). Let us assume that these durations are random variables, say, independent of each other, distributed in given segments $\Delta_\gamma = [\underline{\zeta}_\gamma, \overline{\zeta}_\gamma]$. To avoid pathologies, assume also that $\underline{\zeta}_\gamma \geq \overline{x}_\gamma$ for every γ ("you cannot make the duration negative"). Now (5.2.5) becomes an uncertain LO program with uncertainties affecting only the right hand sides of the constraints. A natural way to "immunize" the solutions to the problem against data uncertainty is to pass to the usual RC of the problem — to think of both t_γ and x_γ as of variables with values to be chosen in advance in such a way that the constraints in (5.2.4) are satisfied for all values of the data ζ_γ from the uncertainty set. With our model of the latter set the RC is nothing but the "worst instance" of our uncertain problem, the one where ζ_γ are set to their maximum possible values $\overline{\zeta}_\gamma$. For large PERT graphs, such an approach is very conservative: why should we care about the highly improbable case where all the normal durations — independent random variables! — are simultaneously at their worst-case values? Note that even taking into account that the normal durations are random and replacing the uncertain constraints in (5.2.5) by their chance constrained versions, we essentially do not reduce the conservatism. Indeed, every one of randomly perturbed constraints in (5.2.5) contains a *single* random perturbation, so that we cannot hope that random perturbations of a constraint will to some extent cancel each other. As a result, to require the validity of every uncertain constraint with probability 0.9 or 0.99 is the same as to require its validity "in the worst case" with just slightly reduced maximal normal durations of the activities.

A much more promising approach is to try to adjust our decisions "on line." Indeed, we are speaking about a process that evolves in time, with "actual decisions" represented by variables x_γ and t_ν's being the analysis variables. Assuming that the decision on x_γ can be postponed till the event μ_γ (the earliest time when the activity γ can be started) takes place, at that time we already know the actual durations of the activities terminated before the event μ_γ, we could then adjust our decision on x_γ in accordance with this information. The difficulty is that *we do not know in advance what will be the actual time precedences between the events — these precedences depend on our decisions and on the actual values of the uncertain data.* For example, in the situation described by figure 5.1, we, in general, cannot know in advance which one of the events B, C will precede the other one in time. As a result, in our present situation, in sharp contrast to the situation of Example 5.2, an attempt to fully utilize the possibilities to adjust the decisions to the actual values of the data results in an extremely complicated problem, where not only the decisions themselves, but the very information base of the decisions become dependent on the uncertain data and our policy. However, we could stick to something in-between "no adjustability at all" and "as much adjustability as possible." Specifically, we definitely know that if a pair of activities γ', γ are linked by a logical precedence, so that there exists an oriented route in the graph that starts with γ' and ends with γ, then the actual duration of γ' will be known before γ can start. Consequently, we can take, as the information base of an activity γ, the collection $\zeta^\gamma = \{\zeta_{\gamma'} : \gamma' \in \Gamma_-(\gamma)\}$, where $\Gamma_-(\gamma)$ is the set of all activities that logically precede the activity γ. In favorable circumstances, such an approach could reduce significantly the price of robustness as compared to the non-adjustable RC. Indeed, when plugging into the randomly perturbed constraints of (5.2.5) instead of constants x_γ functions $X_\gamma(\zeta^\gamma)$, and requiring from the resulting inequalities to be valid with probability $1-\epsilon$, we end up with a system of chance constraints such that some of them (in good cases, even most of them) involve many independent random perturbations each. When the functions $X_\gamma(\zeta^\gamma)$ are regular enough, (e.g., are affine), we can hope that the numerous independent perturbations affecting a chance constraint will to some extent cancel each other, and consequently, the resulting system of chance constraints will be significantly less conservative than the one corresponding to non-adjustable decisions.

5.2.2 Good News on the ARC

Passing from a trivial information base to a nontrivial one — passing from robust optimal *data-independent decisions* to robust optimal *data-based decision rules* can indeed dramatically reduce the associated robust optimal value.

Example 5.4 Consider the toy uncertain LO problem

$$\left\{ \min_x \left\{ x_1 : \begin{array}{ll} x_2 \geq \frac{1}{2}\zeta x_1 + 1 & (a_\zeta) \\ x_1 \geq (2-\zeta)x_2 & (b_\zeta) \\ x_1, x_2 \geq 0 & (c_\zeta) \end{array} \right\} : 0 \leq \zeta \leq \rho \right\},$$

where $\rho \in (0,1)$ is a parameter (uncertainty level). Let us compare the optimal value of its non-adjustable RC (where both x_1 and x_2 must be independent of ζ) with the optimal value of the ARC where x_1 still is assumed to be independent of ζ ($P_1\zeta \equiv 0$) but x_2 is allowed to depend on ζ ($P_2\zeta \equiv \zeta$).

A feasible solution (x_1, x_2) of the RC should remain feasible for the constraint (a_ζ) when $\zeta = \rho$, meaning that $x_2 \geq \frac{\rho}{2}x_1 + 1$, and should remain feasible for the constraint (b_ζ) when $\zeta = 0$, meaning that $x_1 \geq 2x_2$. The two resulting inequalities imply that $x_1 \geq \rho x_1 + 2$, whence $x_1 \geq \frac{2}{1-\rho}$. Thus, Opt(RC)$\geq \frac{2}{1-\rho}$, whence Opt(RC)$\to \infty$ as $\rho \to 1 - 0$.

Now let us solve the ARC. Given $x_1 \geq 0$ and $\zeta \in [0, \rho]$, it is immediately seen that x_1 can be extended, by properly chosen x_2, to a feasible solution of (a_ζ) through (c_ζ) if and only if the pair $(x_1, x_2 = \frac{1}{2}\zeta x_1 + 1)$ is feasible for (a_ζ) through (c_ζ), that is, if and only if $x_1 \geq (2-\zeta)\left[\frac{1}{2}\zeta x_1 + 1\right]$ whenever $1 \leq \zeta \leq \rho$. The latter relation holds true when $x_1 = 4$ and $\rho \leq 1$ (since $(2-\zeta)\zeta \leq 1$ for $0 \leq \zeta \leq 2$). Thus, Opt(ARC)≤ 4, and the difference between Opt(RC) and Opt(ARC) and the ratio Opt(RC)/Opt(ARC) go to ∞ as $\rho \to 1 - 0$.

5.2.3 Bad News on the ARC

Unfortunately, from the computational viewpoint the ARC of an uncertain problem more often than not is wishful thinking rather than an actual tool. The reason comes from the fact that *the ARC is typically severely computationally intractable.* Indeed, (5.2.2) is an *infinite-dimensional problem*, where one wants to optimize over *functions* — decision rules — rather than vectors, and these functions, in general, depend on many real variables. It is unclear even how to *represent* a general-type candidate decision rule — a general-type multivariate function — in a computer. Seemingly the only option here is sticking to a chosen in advance *parametric family* of decision rules, like piece-wise constant/linear/quadratic functions of $P_j\zeta$ with simple domains of the pieces (say, boxes). With this approach, a candidate decision rule is identified by the vector of values of the associated parameters, and the ARC becomes a finite-dimensional problem, the parameters being our new decision variables. This approach is indeed possible and in fact will be the focus of what follows. However, it should be clear from the very beginning that if the parametric family in question is "rich enough" to allow for good approximation of "truly optimal" decision rules (think of polynomial splines of high degree as approximations to "not too rapidly varying" general-type multivariate functions), the number of parameters involved should be astronomically large, unless the dimension of ζ is really small, like 1 — 3 (think of how many coefficients there are in a single algebraic polynomial of degree 10 with 20 variables). Thus, aside of "really low dimensional" cases, "rich" general-purpose parametric families of decision rules are for all practical purposes as intractable as non-parametric families. In other words, when the dimension L of ζ is not too small, tractability of parametric families of decision rules is something opposite to their "approximation abilities," and sticking to tractable parametric families, we lose control of how far the optimal value of the "parametric" ARC is away from the optimal value of the "true" infinite-dimensional ARC. The only exception here seems to be the case when we are smart enough to utilize our knowledge of the structure of instances of the uncertain problem in question in order to identify the optimal decision rules up to a moderate number of parameters. *If* we indeed are that smart and *if* the parameters in question can be further identified numerically in a computationally efficient fashion, we indeed can end up with an optimal solution to the "true" ARC. Unfortunately, the two "if's" in the previous sentence are big if's indeed — to the best of our knowledge, the only *generic* situation when these conditions are satisfied is one treated the Dynamic Programming techniques. It seems that these techniques form the only component in the existing "optimization toolbox" that could be used to process the ARC numerically, at least when approximations of a provably high quality are sought. Unfortunately, the Dynamic Programming techniques are very "fragile" — they require instances of a very specific structure, suffer from "curse of dimensionality," etc. The bottom line, in our opinion, is that *aside of situations where Dynamic Programming is computationally efficient*, (which is an exception rather than a rule), *the only hopefully computationally tractable approach to optimizing over decision rules is to stick to their simple parametric families*, even at the price of giving up full control over the losses in optimality that can be incurred by such a simplification.

Before moving to an in-depth investigation of (a version of) the just outlined "simple approximation" approach to adjustable robust decision-making, it is worth pointing out two situations when no simple approximations are necessary, since the situations in question are very simple from the very beginning.

Simple case I: fixed recourse and scenario-generated uncertainty set

Consider an uncertain *conic* problem

$$\mathcal{P} = \left\{ \min_x \left\{ c_\zeta^T x + d_\zeta : A_\zeta x + b_\zeta \in \mathbf{K} \right\} : \zeta \in \mathcal{Z} \right\} \tag{5.2.6}$$

($A_\zeta, b_\zeta, c_\zeta, d_\zeta$ are affine in ζ, $\mathbf{K}$ is a computationally tractable convex cone) and assume that

1. $\mathcal{Z}$ is a scenario-generated uncertainty set, that is, a set given as a convex hull of finitely many "scenarios" ζ^s, $1 \le s \le S$;
2. The information base ensures that every variable x_j either is non-adjustable ($P_j = 0$), or is fully adjustable ($P_j = I$);
3. We are in the situation of *fixed recourse*, that is, for every adjustable variable x_j (one with $P_j \neq 0$), all its coefficients in the objective and the left hand side of the constraint are certain, (i.e., are independent of ζ).

W.l.o.g. we can assume that $x = [u; v]$, where the u variables are non-adjustable, and the v variables are fully adjustable; under fixed recourse, our uncertain problem can be written down as

$$\mathcal{P} = \left\{ \min_{u,v} \left\{ p_\zeta^T u + q^T v + d_\zeta : P_\zeta u + Qv + r_\zeta \in \mathbf{K} \right\} : \zeta \in \mathrm{Conv}\{\zeta^1, ..., \zeta^S\} \right\}$$

($p_\zeta, d_\zeta, P_\zeta, r_\zeta$ are affine in ζ). An immediate observation is that:

Theorem 5.1 *Under assumptions 1 – 3, the ARC of the uncertain problem $\mathcal{P}$ is equivalent to the computationally tractable conic problem*

$$Opt = \min_{t,u,\{v^s\}_{s=1}^S} \left\{ t : p_{\zeta^s} u + q^T v^s + d_{\zeta^s} \le t,\ P_{\zeta^s} u + Qv^s + r_{\zeta^s} \in \mathbf{K} \right\}. \tag{5.2.7}$$

Specifically, the optimal values in the latter problem and in the ARC of $\mathcal{P}$ are equal. Moreover, if $\bar{t}, \bar{u}, \{\bar{v}^s\}_{s=1}^S$ is a feasible solution to (5.2.7)*, then the pair $\bar{t}, \bar{u}$ augmented by the decision rule for the adjustable variables:*

$$v = \bar{V}(\zeta) = \sum_{s=1}^S \lambda_s(\zeta) \bar{v}^s$$

form a feasible solution to the ARC. Here $\lambda(\zeta)$ is an arbitrary nonnegative vector with the unit sum of entries such that

$$\zeta = \sum_{s=1}^S \lambda_s(\zeta) \zeta^S. \tag{5.2.8}$$

Proof. Observe first that $\lambda(\zeta)$ is well-defined for every $\zeta \in \mathcal{Z}$ due to $\mathcal{Z} = \mathrm{Conv}\{\zeta^1, ..., \zeta^S\}$. Further, if $\bar{t}, \bar{u}, \{\bar{v}^s\}$ is a feasible solution of (5.2.7) and $\bar{V}(\zeta)$ is as defined above, then for every $\zeta \in \mathcal{Z}$ the following implications hold true:

$$\begin{array}{l}
\bar{t} \geq p_{\zeta^s}\bar{u} + q^T\bar{v}^s + d_{\zeta^s}\,\forall s \Rightarrow \bar{t} \geq \sum_s \lambda_s(\zeta)\left[p_{\zeta^s}^T\bar{u} + q^T\bar{v}^s + d_{\zeta^s}\right] \\
= p_\zeta^T\bar{u} + q^T\bar{V}(\zeta) + d_\zeta, \\
\mathbf{K} \ni P_{\zeta^s}\bar{u} + Q\bar{v}^s + r_{\zeta^s}\,\forall s \Rightarrow \mathbf{K} \ni \sum_s \lambda_s(\zeta)\left[P_{\zeta^s}\bar{u} + Q\bar{v}^s + r_{\zeta^s}\right] \\
= P_\zeta\bar{u} + Q\bar{V}(\zeta) + r_\zeta
\end{array}$$

(recall that $p_\zeta, ..., r_\zeta$ are affine in ζ). We see that $(\bar{t}, \bar{u}, \bar{V}(\cdot))$ is indeed a feasible solution to the ARC

$$\min_{t,u,V(\cdot)} \left\{t : p_\zeta^T u + q^T V(\zeta) + d_\zeta \leq t, P_\zeta u + QV(\zeta) + r_\zeta \in \mathbf{K}\,\forall \zeta \in \mathcal{Z}\right\}$$

of $\mathcal{P}$. As a result, the optimal value of the latter problem is $\leq$ Opt. It remains to verify that the optimal value of the ARC and Opt are equal. We already know that the first quantity is $\leq$ the second one. To prove the opposite inequality, note that if $(t, u, V(\cdot))$ is feasible for the ARC, then clearly $(t, u, \{v^s = V(\zeta^s)\})$ is feasible for (5.2.7). □

The outlined result shares the same shortcoming as Theorem 3.1 from section 3.2.1: scenario-generated uncertainty sets are usually too "small" to be of much interest, unless the number L of scenarios is impractically large. It is also worth noticing that the assumption of fixed recourse is essential: it is easy to show (see [14]) that without it, the ARC may become intractable.

Simple case II: uncertain LO with constraint-wise uncertainty

Consider an uncertain LO problem

$$\mathcal{P} = \left\{\min_x \left\{c_\zeta^T x + d_\zeta : a_{i\zeta}^T x \leq b_{i\zeta},\, i = 1, ..., m\right\} : \zeta \in \mathcal{Z}\right\}, \tag{5.2.9}$$

where, as always, $c_\zeta, d_\zeta, a_{i\zeta}, b_{i\zeta}$ are affine in ζ. Assume that

1. The uncertainty is constraint-wise: ζ can be split into blocks $\zeta = [\zeta^0; ...; \zeta^m]$ in such a way that the data of the objective depend solely on ζ^0, the data of the i-th constraint depend solely on ζ^i, and the uncertainty set $\mathcal{Z}$ is the direct product of convex compact sets $\mathcal{Z}_0, \mathcal{Z}_1, ..., \mathcal{Z}_m$ in the spaces of $\zeta^0, ..., \zeta^m$;

2. One can point out a convex compact set $\mathcal{X}$ in the space of x variables such that whenever $\zeta \in \mathcal{Z}$ and x is feasible for the instance of $\mathcal{P}$ with the data ζ, one has $x \in \mathcal{X}$.
   The validity of the latter, purely technical, assumption can be guaranteed, e.g., when the constraints of the uncertain problem contain (certain) finite upper and lower bounds on every one of the decision variables. The latter assumption, for all practical purposes, is non-restrictive.

Our goal is to prove the following

Theorem 5.2 *Under the just outlined assumptions i) and ii), the ARC of* (5.2.9) *is equivalent to its usual RC (no adjustable variables): both ARC and RC have equal optimal values.*

Proof. All we need is to prove that the optimal value in the ARC is $\geq$ the one of the RC. When achieving this goal, we can assume w.l.o.g. that all decision variables are *fully adjustable* — are allowed to depend on the entire vector ζ. The "fully adjustable" ARC of (5.2.9) reads

$$\begin{aligned}\mathrm{Opt}(\mathrm{ARC}) &= \min_{X(\cdot),t}\left\{t: \begin{array}{l} c_{\zeta^0}^T X(\zeta) + d_{\zeta^0} - t \leq 0 \\ a_{i\zeta^i}^T X(\zeta) - b_{i\zeta^i} \leq 0, 1 \leq i \leq m \end{array}\right. \\ &\qquad\qquad\qquad \left.\forall(\zeta \in \mathcal{Z}_0 \times ... \times \mathcal{Z}_m)\right\} \\ &= \inf\Big\{t : \forall(\zeta \in \mathcal{Z}_0 \times ... \times \mathcal{Z}_m)\exists x \in \mathcal{X} : \\ &\qquad\qquad \alpha_{i,\zeta^i}^T x - \beta_i t + \gamma_{i,\zeta^i} \leq 0,\ 0 \leq i \leq m\Big\},\end{aligned} \tag{5.2.10}$$

(the restriction $x \in \mathcal{X}$ can be added due to assumption i)), while the RC is the problem

$$\mathrm{Opt}(\mathrm{RC}) = \inf\left\{t : \exists x \in \mathcal{X} : \alpha_{i\zeta^i}^T x - \beta_i t + \gamma_{i\zeta^i} \leq 0 \forall(\zeta \in \mathcal{Z}_0 \times ... \times \mathcal{Z}_m)\right\}; \tag{5.2.11}$$

here $\alpha_{i\zeta^i}$, $\gamma_{i\zeta^i}$ are affine in ζ^i and $\beta_i \geq 0$.

In order to prove that $\mathrm{Opt}(\mathrm{ARC}) \geq \mathrm{Opt}(\mathrm{RC})$, it suffices to consider the case when $\mathrm{Opt}(\mathrm{ARC}) < \infty$ and to show that whenever a real $\bar{t}$ is $> \mathrm{Opt}(\mathrm{ARC})$, we have $\bar{t} \geq \mathrm{Opt}(\mathrm{RC})$. Looking at (5.2.11), we see that to this end it suffices to lead to a contradiction the statement that for some $\bar{t} > \mathrm{Opt}(\mathrm{ARC})$ one has

$$\forall x \in \mathcal{X} \exists (i = i_x \in \{0, 1, ..., m\}, \zeta^i = \zeta_x^{i_x} \in \mathcal{Z}_{i_x}) : \alpha_{i_x \zeta_x^{i_x}}^T x - \beta_{i_x} \bar{t} + \gamma_{i_x \zeta_x^{i_x}} > 0. \tag{5.2.12}$$

Assume that $\bar{t} > \mathrm{Opt}(\mathrm{ARC})$ and that (5.2.12) holds. For every $x \in \mathcal{X}$, the inequality

$$\alpha_{i_x \zeta_x^{i_x}}^T y - \beta_{i_x} \bar{t} + \gamma_{i_x \zeta_x^{i_x}} > 0$$

is valid when $y = x$; therefore, for every $x \in \mathcal{X}$ there exist $\epsilon_x > 0$ and a neighborhood U_x of x such that

$$\forall y \in U_x : \alpha_{i_x \zeta_x^{i_x}}^T y - \beta_{i_x} \bar{t} + \gamma_{i_x \zeta_x^{i_x}} \geq \epsilon_x.$$

Since $\mathcal{X}$ is a compact set, we can find finitely many points $x^1, ..., x^N$ such that $\mathcal{X} \subset \bigcup_{j=1}^N U_{x^j}$. Setting $\epsilon = \min_j \epsilon_{x^j}$, $i[j] = i_{x^j}$, $\zeta[j] = \zeta_{x^j}^{i_{x^j}} \in \mathcal{Z}_{i[j]}$, and

$$f_j(y) = \alpha_{i[j],\zeta[j]}^T y - \beta_{i[j]} \bar{t} + \gamma_{i[j],\zeta[j]},$$

we end up with N affine functions of y such that

$$\max_{1 \leq j \leq N} f_j(y) \geq \epsilon > 0\ \forall y \in \mathcal{X}.$$

Since $\mathcal{X}$ is a convex compact set and $f_j(\cdot)$ are affine (and thus convex and continuous) functions, the latter relation, by well-known facts from Convex Analysis (namely, the von Neumann Lemma), implies that there exists a collection of nonnegative weights λ_j with $\sum_j \lambda_j = 1$ such that

$$f(y) \equiv \sum_{j=1}^N \lambda_j f_j(y) \geq \epsilon\, \forall y \in \mathcal{X}. \tag{5.2.13}$$

Now let

$$\begin{array}{rcl}\omega_i &=& \sum_{j:i[j]=i}\lambda_j,\ i=0,1,...,m;\\ \bar{\zeta}^i &=& \left\{\begin{array}{ll}\sum_{j:i[j]=i}\frac{\lambda_j}{\omega_i}\zeta[j], & \omega_i>0\\ \text{a point from } \mathcal{Z}_i, & \omega_i=0\end{array}\right.,\\ \bar{\zeta} &=& [\bar{\zeta}^0;...;\bar{\zeta}^m].\end{array}$$

Due to its origin, every one of the vectors $\bar{\zeta}^i$ is a convex combination of points from $\mathcal{Z}_i$ and as such belongs to $\mathcal{Z}_i$, since the latter set is convex. Since the uncertainty is constraint-wise, we conclude that $\bar{\zeta}\in\mathcal{Z}$. Since $\bar{t}>\text{Opt(ARC)}$, we conclude from (5.2.10) that there exists $\bar{x}\in\mathcal{X}$ such that the inequalities

$$\alpha_{i\bar{\zeta}^i}^T\bar{x}-\beta_i\bar{t}+\gamma_{i\bar{\zeta}^i}\leq 0$$

hold true for every i, $0\leq i\leq m$. Taking a weighted sum of these inequalities, the weights being ω_i, we get

$$\sum_{i:\omega_i>0}\omega_i[\alpha_{i\bar{\zeta}^i}^T\bar{x}-\beta_i\bar{t}+\gamma_{i\bar{\zeta}^i}]\leq 0. \tag{5.2.14}$$

At the same time, by construction of $\bar{\zeta}^i$ and due to the fact that $\alpha_{i\zeta^i}$, $\gamma_{i\zeta^i}$ are affine in ζ^i, for every i with $\omega_i>0$ we have

$$[\alpha_{i\bar{\zeta}^i}^T\bar{x}-\beta_i\bar{t}+\gamma_{i\bar{\zeta}^i}]=\sum_{j:i[j]=i}\frac{\lambda_j}{\omega_i}f_j(\bar{x}),$$

so that (5.2.14) reads

$$\sum_{j=1}^N\lambda_j f_j(\bar{x})\leq 0,$$

which is impossible due to (5.2.13) and to $\bar{x}\in\mathcal{X}$. We have arrived at the desired contradiction. □

5.3 Affinely Adjustable Robust Counterparts

We are about to investigate in-depth a specific version of the "parametric decision rules" approach we have outlined previously. At this point, we prefer to come back from general-type uncertain problem (5.1.1) to affinely perturbed uncertain conic problem

$$\mathcal{C}=\left\{\min_{x\in\mathbb{R}^n}\left\{c_\zeta^Tx+d_\zeta : A_\zeta x+b_\zeta\in\mathbf{K}\right\}:\zeta\in\mathcal{Z}\right\}, \tag{5.3.1}$$

where $c_\zeta, d_\zeta, A_\zeta, b_\zeta$ are affine in ζ, $\mathbf{K}$ is a "nice" cone (direct product of nonnegative rays/Lorentz cones/semidefinite cones, corresponding to uncertain LP/CQP/SDP, respectively), and $\mathcal{Z}$ is a convex compact uncertainty set given by a strictly feasible SDP representation

$$\mathcal{Z}=\left\{\zeta\in\mathbb{R}^L:\exists u:\mathcal{P}(\zeta,u)\succeq 0\right\},$$

where $\mathcal{P}$ is affine in $[\zeta;u]$. Assume that along with the problem, we are given an information base $\{P_j\}_{j=1}^n$ for it; here P_j are $m_j\times n$ matrices. To save words (and without risk of ambiguity), we shall call such a pair "uncertain problem $\mathcal{C}$, information base" merely an *uncertain conic* *problem.* Our course of action is to restrict the ARC of the problem to a specific parametric family of decision rules, namely, the *affine* ones:

$$x_j = X_j(P_j\zeta) = p_j + q_j^T P_j\zeta,\ j = 1, ..., n. \tag{5.3.2}$$

The resulting restricted version of the ARC of (5.3.1), which we call the *Affinely Adjustable Robust Counterpart* (AARC), is the semi-infinite optimization program

$$\min_{t,\{p_j,q_j\}_{j=1}^n} \left\{ t : \left. \begin{array}{l} \sum_{j=1}^n c_\zeta^j [p_j + q_j^T P_j\zeta] + d_\zeta - t \leq 0 \\ \sum_{j=1}^n A_\zeta^j [p_j + q_j^T P_j\zeta] + b_\zeta \in \mathbf{K} \end{array} \right\} \forall \zeta \in \mathcal{Z} \right\}, \tag{5.3.3}$$

where c_ζ^j is j-th entry in c_ζ, and A_ζ^j is j-th column of A_ζ. Note that the variables in this problem are t and the coefficients p_j, q_j of the affine decision rules (5.3.2). As such, these variables do *not* specify uniquely the actual decisions x_j; these decisions are uniquely defined by these coefficients *and* the corresponding portions $P_j\zeta$ of the true data once the latter become known.

5.3.1 Tractability of the AARC

The rationale for focusing on affine decision rules rather than on other parametric families is that *there exists at least one important case when the AARC of an uncertain conic problem is, essentially, as tractable as the RC of the problem.* The "important case" in question is the one of *fixed recourse* and is defined as follows:

Definition 5.1 *Consider an uncertain conic problem* (5.3.1) *augmented by an information base* $\{P_j\}_{j=1}^n$*. We say that this pair is with fixed recourse, if the coefficients of every adjustable, (i.e., with* $P_j \neq 0$*), variable* x_j *are certain:*

$$\forall (j : P_j \neq 0) : \text{ both } c_\zeta^j \text{ and } A_\zeta^j \text{ are independent of } \zeta.$$

For example, both Examples 5.1 (Inventory) and 5.2 (Project Management) are uncertain problems with fixed recourse.

An immediate observation is as follows:

> **(!)** *In the case of fixed recourse, the AARC, similarly to the RC, is a semi-infinite conic problem — it is the problem*
>
> $$\min_{t,y=\{p_j,q_j\}} \left\{ t : \left. \begin{array}{l} \widehat{c}_\zeta^T y + d_\zeta \leq t \\ \widehat{A}_\zeta y + b_\zeta \in \mathbf{K} \end{array} \right\} \forall \zeta \in \mathcal{Z} \right\}, \tag{5.3.4}$$
>
> *with* $\widehat{c}_\zeta, d_\zeta, \widehat{A}_\zeta, b_\zeta$ *affine in* ζ*:*
>
> $$\begin{array}{rcl} \widehat{c}_\zeta^T y &=& \sum_j c_\zeta^j [p_j + q_j^T P_j\zeta] \\ \widehat{A}_\zeta y &=& \sum_j A_\zeta^j [p_j + q_j^T P_j\zeta]. \end{array} \qquad [y = \{[p_j, q_j]\}_{j=1}^n]$$

Note that it is exactly fixed recourse that makes $\widehat{c}_\zeta, \widehat{A}_\zeta$ affine in ζ; without this assumption, these entities are quadratic in ζ.

As far as the tractability issues are concerned, observation (!) is *the* main argument in favor of affine decision rules, *provided we are in the situation of fixed recourse.* Indeed, in the latter situation the AARC is a semi-infinite conic problem, and we can apply to it all the results of previous lectures related to tractable reformulations/tight safe tractable approximations of semi-infinite conic problems. Note that many of these results, while imposing certain restrictions on the geometries of the uncertainty set and the cone $\mathbf{K}$, require from the objective (if it is uncertain) and the left hand sides of the uncertain constraints nothing more than bi-affinity in the decision variables and in the uncertain data. *Whenever this is the case, the "tractability status" of the AARC is not worse than the one of the usual RC.* In particular, *in the case of fixed recourse* we can:

1. Convert the AARC of an uncertain LO problem into an explicit efficiently solvable "well-structured" convex program (see Theorem 1.1).

2. Process efficiently the AARC of an uncertain conic quadratic problem with (common to all uncertain constraints) simple ellipsoidal uncertainty (see section 3.2.5).

3. Use a tight safe tractable approximation of an uncertain problem with linear objective and convex quadratic constraints with (common for all uncertain constraints) ∩-ellipsoidal uncertainty (see section 3.3.2): whenever $\mathcal{Z}$ is the intersection of M ellipsoids centered at the origin, the problem admits a safe tractable approximation tight within the factor $O(1)\sqrt{\ln(M)}$ (see Theorem 3.11).

The reader should be aware, however, that the AARC, in contrast to the usual RC, is *not* a constraint-wise construction, since when passing to the coefficients of affine decision rules as our new decision variables, the portion of the uncertain data affecting a particular constraint can change when allowing the original decision variables entering the constraint to depend on the uncertain data not affecting the constraint directly. This is where the words "common" in the second and the third of the above statements comes from. For example, the RC of an uncertain conic quadratic problem with the constraints of the form

$$\|A^i_\zeta x + b^i_\zeta\|_2 \leq x^T c^i_\zeta + d^i_\zeta,\ i = 1, ..., m,$$

is computationally tractable, provided that the projection $\mathcal{Z}_i$ of the overall uncertainty set $\mathcal{Z}$ onto the subspace of data perturbations of i-th constraint is an ellipsoid (section 3.2.5). To get a similar result for the AARC, we need *the overall uncertainty set $\mathcal{Z}$ itself* to be an ellipsoid, since otherwise the projection of $\mathcal{Z}$ on the data of the "AARC counterparts" of original uncertain constraints can be different from ellipsoids. The bottom line is that the claim that with fixed recourse, the AARC of an uncertain problem is "as tractable" as its RC should be understood with some caution. This, however, is not a big deal, since the "recipe" is already here: *Under the assumption of fixed recourse, the AARC is a semi-infinite conic problem, and in order to process it computationally, we can use all the machinery developed in the previous lectures. If this machinery allows for tractable reformulation/tight safe tractable approximation of the problem, fine, otherwise too bad for us."* Recall that there exists at least one really important case when everything is fine — this is the case of uncertain LO problem with fixed recourse.

It should be added that when processing the AARC in the case of fixed recourse, we can enjoy all the results on safe tractable approximations of chance constrained affinely perturbed scalar, conic quadratic and linear matrix inequalities developed in previous lectures. Recall that these results imposed certain restrictions on the distribution of ζ (like independence of $\zeta_1, ..., \zeta_L$), but never required more than affinity of the bodies of the constraints w.r.t. ζ, so that these results work equally well in the cases of RC and AARC.

Last, but not least, the concept of an Affinely Adjustable Robust Counterpart can be straightforwardly "upgraded" to the one of Affinely Adjustable *Globalized* Robust Counterpart. We have no doubts that a reader can carry out such an "upgrade" on his/her own and understands that in the case of fixed recourse, the above "recipe" is equally applicable to the AARC and the AAGRC.

5.3.2 Is Affinity an Actual Restriction?

Passing from *arbitrary* decision rules to *affine* ones seems to be a dramatic simplification. On a closer inspection, the simplification is not as severe as it looks, or, better said, the "dramatics" is not exactly where it is seen at first glance. Indeed, assume that we would like to use decision rules that are quadratic in $P_j\zeta$ rather than linear. Are we supposed to introduce a special notion of a "Quadratically Adjustable Robust Counterpart"? The answer is negative. All we need is to augment the data vector $\zeta = [\zeta_1; ...; \zeta_L]$ by extra entries — the pairwise products $\zeta_i\zeta_j$ of the original entries — and to treat the resulting "extended" vector $\widehat{\zeta} = \widehat{\zeta}[\zeta]$ as our new uncertain data. With this, the decision rules that are quadratic in $P_j\zeta$ become *affine* in $\widehat{P}_j\widehat{\zeta}[\zeta]$, where $\widehat{P}_j$ is a matrix readily given by P_j. More generally, assume that we want to use decision rules of the form

$$X_j(\zeta) = p_j + q_j^T \widehat{P}_j\widehat{\zeta}[\zeta], \tag{5.3.5}$$

where $p_j \in \mathbb{R}, q_j \in \mathbb{R}^{m_j}$ are "free parameters," (which can be restricted to reside in a given convex set), $\widehat{P}_j$ are given $m_j \times D$ matrices and

$$\zeta \mapsto \widehat{\zeta}[\zeta] : \mathbb{R}^L \to \mathbb{R}^D$$

is a given, possibly nonlinear, mapping. Here again we can pass from the original data vector ζ to the data vector $\widehat{\zeta}[\zeta]$, thus making the desired decision rules (5.3.5) merely affine in the "portions" $\widehat{P}_j\widehat{\zeta}$ of the new data vector. We see that when allowing for a seemingly harmless redefinition of the data vector, affine decision rules become as powerful as arbitrary affinely parameterized parametric families of decision rules. This latter class is really huge and, for all practical purposes, is as rich as the class of *all* decision rules. Does it mean that the concept of AARC is basically as flexible as the one of ARC? Unfortunately, the answer is negative, and the reason for the negative answer comes not from potential difficulties with extremely complicated nonlinear transformations $\zeta \mapsto \widehat{\zeta}[\zeta]$ and/or "astronomically large" dimension D of the transformed data vector. The difficulty arises already when the transformation is pretty simple, as is the case, e.g., when the coordinates in $\widehat{\zeta}[\zeta]$ are just the entries of ζ and the pairwise products of these entries. Here is where the difficulty arises. Assume that we are speaking about a single uncertain affinely perturbed scalar linear constraint, allow for quadratic dependence of the original decision variables on the data and pass to the associated adjustable robust counterpart of the constraint. As it was just explained, this counterpart is nothing but a semi-infinite scalar inequality

$$\forall(\widehat{\zeta} \in \mathcal{U}) : a_{0,\widehat{\zeta}} + \sum_{j=1}^{J} a_{j,\widehat{\zeta}} y_j \leq 0$$

where $a_{j,\widehat{\zeta}}$ are affine in $\widehat{\zeta}$, the entries in $\widehat{\zeta} = \widehat{\zeta}[\zeta]$ are the entries in ζ and their pairwise products, $\mathcal{U}$ is the image of the "true" uncertainty set $\mathcal{Z}$ under the *nonlinear* mapping $\zeta \to \widehat{\zeta}[\zeta]$, and y_j are our new decision variables (the coefficients of the quadratic decision rules). While the body of the constraint in question is bi-affine in y and in $\widehat{\zeta}$, this semi-infinite constraint can well be intractable, since the *uncertainty set $\mathcal{U}$ may happen to be intractable, even when $\mathcal{Z}$ is tractable.* Indeed, the tractability of a semi-infinite bi-affine scalar constraint

$$\forall(u \in \mathcal{U}) : f(y, u) \leq 0$$

heavily depends on whether the underlying uncertainty set $\mathcal{U}$ is convex and computationally tractable. When it is the case, we can, modulo minor technical assumptions, solve efficiently the *Analysis problem* of checking whether a given candidate solution y is feasible for the constraint — to this end, it suffices to maximize the affine function $f(y, \cdot)$ over the computationally tractable convex set $\mathcal{U}$. This, under minor technical assumptions, can be done efficiently. The latter fact, in turn, implies (again modulo minor technical assumptions) that we can optimize efficiently linear/convex objectives under the constraints with the above features, and this is basically all we need. The situation changes dramatically when the uncertainty set $\mathcal{U}$ is *not* a convex computationally tractable set. By itself, the convexity of $\mathcal{U}$ costs nothing: since f is bi-affine, the feasible set of the semi-infinite constraint in question remains intact when we replace $\mathcal{U}$ with its convex hull $\widehat{\mathcal{Z}}$. The actual difficulty is that the convex hull $\widehat{\mathcal{Z}}$ of the set $\mathcal{U}$ can be computationally intractable. In the situation we are interested in — the one where $\widehat{\mathcal{Z}} = \text{Conv}\mathcal{U}$ and $\mathcal{U}$ is the image of a computationally tractable convex set $\mathcal{Z}$ under a *nonlinear* transformation $\zeta \mapsto \widehat{\zeta}[\zeta]$, $\widehat{\mathcal{Z}}$ can be computationally intractable already for pretty simple $\mathcal{Z}$ and nonlinear mappings $\zeta \mapsto \widehat{\zeta}[\zeta]$. It happens, e.g., when $\mathcal{Z}$ is the unit box $\|\zeta\|_\infty \leq 1$ and $\widehat{\zeta}[\zeta]$ is comprised of the entries in ζ and their pairwise products. In other words, the "Quadratically Adjustable Robust Counterpart" of an uncertain linear inequality with interval uncertainty is, in general, computationally intractable.

In spite of the just explained fact that "global linearization" of nonlinear decision rules via nonlinear transformation of the data vector not necessarily leads to tractable adjustable RCs, one should keep in mind this option, since it is important methodologically. Indeed, "global linearization" allows one to "split" the problem of processing the ARC, restricted to decision rules (5.3.5), into two subproblems:

(a) Building a tractable representation (or a tight tractable approximation) of the convex hull $\widehat{\mathcal{Z}}$ of the image $\mathcal{U}$ of the original uncertainty set $\mathcal{Z}$ under the nonlinear mapping $\zeta \mapsto \widehat{\zeta}[\zeta]$ associated with (5.3.5). Note that this problem by itself has nothing to do with adjustable robust counterparts and the like;

(b) Developing a tractable reformulation (or a tight safe tractable approximation) of the *Affinely Adjustable* Robust Counterpart of the uncertain problem in question, with $\widehat{\zeta}$ in the role of the data vector, the tractable convex set, yielded by (a), in the role of the uncertainty set, and the information base given by the matrices $\widehat{P}_j$.

Of course, the resulting two problems are not completely independent: the tractable convex set $\widehat{\mathcal{Z}}$ with which we, upon success, end up when solving (a) should be simple enough to allow for successful processing of (b). Note, however, that this "coupling of problems (a) and (b)" is of no importance when the uncertain problem in question is an LO problem with fixed recourse. Indeed, in this case the AARC of the problem is computationally tractable whatever the uncertainty set as long as it is tractable, therefore every tractable set $\widehat{\mathcal{Z}}$ yielded by processing of problem (a) will do.

Example 5.5 Assume that we want to process an uncertain LO problem

$$\mathcal{C} = \left\{ \min_x \left\{ c_\zeta^T x + d_\zeta : A_\zeta x \geq b_\zeta \right\} : \zeta \in \mathcal{Z} \right\} \quad [c_\zeta, d_\zeta, A_\zeta, b_\zeta : \text{ affine in } \zeta] \tag{5.3.6}$$

with fixed recourse and a tractable convex compact uncertainty set $\mathcal{Z}$, and consider a number of affinely parameterized families of decision rules.

A. *"Genuine" affine decision rules: x_j is affine in $P_j\zeta$.* As we have already seen, the associated ARC — the usual AARC of $\mathcal{C}$ — is computationally tractable.

B. *Piece-wise linear decision rules with fixed breakpoints.* Assume that the mapping $\zeta \mapsto \widehat{\zeta}[\zeta]$ augments the entries of ζ with finitely many entries of the form $\phi_i(\zeta) = \max\left[r_i, s_i^T\zeta\right]$, and the decision rules we intend to use should be affine in $\widehat{P}_j\widehat{\zeta}$, where $\widehat{P}_j$ are given matrices. In order to process the associated ARC in a computationally efficient fashion, all we need is to build a tractable representation of the set $\widehat{\mathcal{Z}} = \mathrm{Conv}\{\widehat{\zeta}[\zeta] : \zeta \in \mathcal{Z}\}$. While this could be difficult in general, there are useful cases when the problem is easy, e.g., the case where

$$\begin{array}{c}\mathcal{Z} = \{\zeta \in \mathbb{R}^L : f_k(\zeta) \le 1, 1 \le k \le K\},\\ \widehat{\zeta}[\zeta] = [\zeta; (\zeta)_+; (\zeta)_-], \text{ with } (\zeta)_- = \max[\zeta, 0_{L\times 1}], (\zeta)_+ = \max[-\zeta, 0_{L\times 1}].\end{array}$$

Here, for vectors u, v, $\max[u, v]$ is taken coordinate-wise, and $f_k(\cdot)$ are lower semicontinuous and *absolutely symmetric* convex functions on $\mathbb{R}^L$, absolute symmetry meaning that $f_k(\zeta) \equiv f_k(\mathrm{abs}(\zeta))$ (abs acts coordinate-wise). (Think about the case when $f_k(\zeta) = \|[\alpha_{k1}\zeta_1; ...; \alpha_{kL}\zeta_L]\|_{p_k}$ with $p_k \in [1, \infty]$.) It is easily seen that if $\mathcal{Z}$ is bounded, then

$$\widehat{\mathcal{Z}} = \left\{\widehat{\zeta} = [\zeta; \zeta^+; \zeta^-] : \begin{array}{ll}(a) & f_k(\zeta^+ + \zeta^-) \le 1,\ 1 \le k \le K\\ (b) & \zeta = \zeta^+ - \zeta^-\\ (c) & \zeta^\pm \ge 0\end{array}\right\}.$$

Indeed, (a) through (c) is a system of convex constraints on vector $\widehat{\zeta} = [\zeta; \zeta^+; \zeta^-]$, and since f_k are lower semicontinuous, the feasible set C of this system is convex and closed; besides, for $[\zeta; \zeta^+; \zeta^-] \in C$ we have $\zeta^+ + \zeta^- \in \mathcal{Z}$; since the latter set is bounded by assumption, the sum $\zeta^+ + \zeta^-$ is bounded uniformly in $\widehat{\zeta} \in C$, whence, by (a) through (c), C is bounded. Thus, C is a closed and bounded convex set. The image $\mathcal{U}$ of the set $\mathcal{Z}$ under the mapping $\zeta \mapsto [\zeta; (\zeta)_+; (\zeta)_-]$ clearly is contained in C, so that the convex hull $\widehat{\mathcal{Z}}$ of $\mathcal{U}$ is contained in C as well. To prove the inverse inclusion, note that since C is a (nonempty) convex compact set, it is the convex hull of the set of its extreme points, and therefore in order to prove that $\widehat{\mathcal{Z}} \supset C$ it suffices to verify that every extreme point $[\zeta; \zeta^+, \zeta^-]$ of C belongs to $\mathcal{U}$. But this is immediate: in an extreme point of C we should have $\min[\zeta_\ell^+, \zeta_\ell^-] = 0$ for every ℓ, since if the opposite were true for some $\ell = \bar{\ell}$, then C would contain a nontrivial segment centered at the point, namely, points obtained from the given one by the "3-entry perturbation" $\zeta_{\bar{\ell}}^+ \mapsto \zeta_{\bar{\ell}}^+ + \delta$, $\zeta_{\bar{\ell}}^- \mapsto \zeta_{\bar{\ell}}^+ - \delta$, $\zeta_{\bar{\ell}} \mapsto \zeta_{\bar{\ell}} + 2\delta$ with small enough $|\delta|$. Thus, every extreme point of C has $\min[\zeta^+, \zeta^-] = 0$, $\zeta = \zeta^+ - \zeta^-$, and a point of this type satisfying (a) clearly belongs to $\mathcal{U}$. □

C. *Separable decision rules.* Assume that $\mathcal{Z}$ is a box: $\mathcal{Z} = \{\zeta : \underline{a} \le \zeta \le \overline{a}\}$, and we are seeking for *separable* decision rules with a prescribed "information base," that is, for the decision rules of the form

$$x_j = \xi_j + \sum_{\ell \in I_j} f_\ell^j(\zeta_\ell),\ j = 1, ..., n, \tag{5.3.7}$$

where the only restriction on functions f_ℓ^j is to belong to given finite-dimensional linear spaces $\mathcal{F}_\ell$ of univariate functions. The sets I_j specify the information base of our decision rules. Some of these sets may be empty, meaning that the associated x_j are non-adjustable decision variables, in full accordance with the standard convention that a sum over an empty set of indices is 0. We consider two specific choices of the spaces $\mathcal{F}_\ell$:

C.1: $\mathcal{F}_\ell$ is comprised of all piecewise linear functions on the real axis with fixed breakpoints $a_{\ell 1} < ... < a_{\ell m}$ (w.l.o.g., assume that $\underline{a}_\ell < a_{\ell 1}$, $a_{\ell m} < \overline{a}_\ell$);

C.2: $\mathcal{F}_\ell$ is comprised of all algebraic polynomials on the axis of degree $\le \kappa$.

Note that what follows works when m in **C.1** and κ in **C.2** depend on ℓ; in order to simplify notation, we do not consider this case explicitly.

C.1: Let us augment every entry ζ_ℓ of ζ with the reals $\zeta_{\ell i}[\zeta_\ell] = \max[\zeta_\ell, a_{\ell i}]$, $i = 1, ..., m$, and let us set $\zeta_{\ell 0}[\zeta_\ell] = \zeta_\ell$. In the case of **C.1**, decision rules (5.3.7) are exactly the rules where x_j is affine in $\{\zeta_{\ell i}[\zeta] : \ell \in I_j\}$; thus, all we need in order to process efficiently the ARC of (5.3.6) restricted to the decision rules in question is a tractable representation of the convex hull of the image $\mathcal{U}$ of $\mathcal{Z}$ under the mapping $\zeta \mapsto \{\zeta_{\ell i}[\zeta]\}_{\ell,i}$. Due to the direct product structure of $\mathcal{Z}$, the set $\mathcal{U}$ is the direct product, over $\ell = 1, ..., d$, of the sets

$$\mathcal{U}_\ell = \{[\zeta_{\ell 0}[\zeta_\ell]; \zeta_{\ell 1}[\zeta_\ell]; ...; \zeta_{\ell m}[\zeta_\ell]] : \underline{a}_\ell \leq \zeta_\ell \leq \overline{a}_\ell\},$$

so that all we need are tractable representations of the convex hulls of the sets $\mathcal{U}_\ell$. The bottom line is, that all we need is a tractable description of a set C of the form

$$C_m = \mathrm{Conv} S_m,\ S_m = \{[s_0; \max[s_0, a_1]; ...; \max[s_0, a_m]] : a_0 \leq s_0 \leq a_{m+1}\},$$

where $a_0 < a_1 < a_2 < ... < a_m < a_{m+1}$ are given. An explicit polyhedral description of the set C_m is given by the following

Lemma 5.1 [3, Lemma 14.3.3] *The convex hull C_m of the set S_m is*

$$C_m = \left\{ [s_0; s_1; ...; s_m] : \left\{ \begin{array}{l} a_0 \leq s_0 \leq a_{m+1} \\ 0 \leq \frac{s_1 - s_0}{a_1 - a_0} \leq \frac{s_2 - s_1}{a_2 - a_1} \leq ... \leq \frac{s_{m+1} - s_m}{a_{m+1} - a_m} \leq 1 \end{array} \right. \right\}, \tag{5.3.8}$$

where $s_{m+1} = a_{m+1}$.

C.2: Similar to the case of **C.1**, in the case of **C.2** all we need in order to process efficiently the ARC of (5.3.6), restricted to decision rules (5.3.7), is a tractable representation of the set

$$C = \mathrm{Conv} S,\ S = \{\widehat{s} = [s; s^2; ...; s^\kappa] : |s| \leq 1\}.$$

(We have assumed w.l.o.g. that $\underline{a}_\ell = -1$, $\overline{a}_\ell = 1$.) Here is the description (originating from [75]):

Lemma 5.2 *The set $C = \mathrm{Conv} S$ admits the explicit semidefinite representation*

$$C = \left\{ \widehat{s} \in \mathbb{R}^\kappa : \exists \lambda = [\lambda_0; ...; \lambda_{2\kappa}] \in \mathbb{R}^{2\kappa+1} : [1; \widehat{s}] = Q^T \lambda, [\lambda_{i+j}]_{i,j=0}^{\kappa} \succeq 0 \right\}, \tag{5.3.9}$$

where the $(2\kappa + 1) \times (\kappa + 1)$ matrix Q is defined as follows: take a polynomial $p(t) = p_0 + p_1 t + ... + p_\kappa t^\kappa$ and convert it into the polynomial $\widehat{p}(t) = (1 + t^2)^\kappa p(2t/(1 + t^2))$. The vector of coefficients of $\widehat{p}$ clearly depends linearly on the vector of coefficients of p, and Q is exactly the matrix of this linear transformation.

Proof. 1^0. Let $P \subset \mathbb{R}^{\kappa+1}$ be the cone of vectors p of coefficients of polynomials $p(t) = p_0 + p_1 t + p_2 t^2 + ... + p_\kappa t^\kappa$ that are nonnegative on $[-1, 1]$, and P_* be the cone dual to P. We claim that

$$C = \{\widehat{s} \in \mathbb{R}^\kappa : [1; \widehat{s}] \in P_*\}. \tag{5.3.10}$$

Indeed, let C' be the right hand side set in (5.3.10). If $\widehat{s} = [s; s^2; ...; s^\kappa] \in S$, then $|s| \leq 1$, so that for every $p \in P$ we have $p^T[1; \widehat{s}] = p(s) \geq 0$. Thus, $[1; \widehat{s}] \in P_*$ and therefore $\widehat{s} \in C'$. Since C' is convex, we arrive at $C \equiv \mathrm{Conv} S \subset C'$. To prove the inverse inclusion, assume that there exists $\widehat{s} \notin C$ such that $z = [1; \widehat{s}] \in P_*$, and let us lead this assumption to a contradiction. Since $\widehat{s}$ is not in C and C is a closed convex set and clearly contains the origin, we can find a vector $q \in \mathbb{R}^\kappa$ such that $q^T \widehat{s} = 1$ and $\max_{r \in C} q^T r \equiv \alpha < 1$, or, which is the same due to $C = \mathrm{Conv} S$, $q^T[s; s^2; ...; s^\kappa] \leq \alpha < 1$ whenever $|s| \leq 1$. Setting $p = [\alpha; -q]$, we see that $p(s) \geq 0$ whenever $|s| \leq 1$, so that $p \in P$ and therefore $\alpha - q^T\widehat{s} = p^T[1;\widehat{s}] \geq 0$, whence $1 = q^T\widehat{s} \leq \alpha < 1$, which is a desired contradiction.

2^0. It remains to verify that the right hand side in (5.3.10) indeed admits representation (5.3.9). We start by deriving a semidefinite representation of the cone P_+ of (vectors of coefficients of) all polynomials $p(s)$ of degree not exceeding 2κ that are nonnegative on the entire axis. The representation is as follows. A $(\kappa+1)\times(\kappa+1)$ symmetric matrix W can be associated with the polynomial of degree $\leq 2\kappa$ given by $p_W(t) = [1;t;t^2;...;t^\kappa]^T W[1;t;t^2;...;t^\kappa]$, and the mapping $\mathcal{A}: W \mapsto p_W$ clearly is linear: $\left(\mathcal{A}[w_{ij}]_{i,j=0}^\kappa\right)_\nu = \sum_{0\leq i\leq \nu} w_{i,\nu-i}$, $0 \leq \nu \leq 2\kappa$. A dyadic matrix $W = ee^T$ "produces" in this way a polynomial that is the square of another polynomial: $\mathcal{A}ee^T = e^2(t)$ and as such is ≥ 0 on the entire axis. Since every matrix $W \succeq 0$ is a sum of dyadic matrices, we conclude that $\mathcal{A}W \in P_+$ whenever $W \succeq 0$. Vice versa, it is well known that every polynomial $p \in P_+$ is the sum of squares of polynomials of degrees $\leq \kappa$, meaning that every $p \in P_+$ is $\mathcal{A}W$ for certain W that is the sum of dyadic matrices and as such is $\succeq 0$. Thus,

$$P_+ = \{p = \mathcal{A}W : W \in \mathbf{S}_+^{\kappa+1}\}.$$

Now, the mapping $t \mapsto 2t/(1+t^2) : \mathbb{R} \to \mathbb{R}$ maps $\mathbb{R}$ onto the segment $[-1,1]$. It follows that *a polynomial p of degree $\leq \kappa$ is ≥ 0 on $[-1,1]$ if and only if the polynomial $\widehat{p}(t) = (1+t^2)^\kappa p(2t/(1+t^2))$ of degree $\leq 2\kappa$ is ≥ 0 on the entire axis*, or, which is the same, $p \in P$ if and only if $Qp \in P_+$. Thus,

$$P = \{p \in \mathbb{R}^{\kappa+1} : \exists W \in \mathbf{S}^{\kappa+1} : W \succeq 0, \mathcal{A}W = Qp\}.$$

Given this semidefinite representation of P, we can immediately obtain a semidefinite representation of P_*. Indeed,

$$\begin{array}{l} q \in P_* \Leftrightarrow 0 \leq \min_{p\in P}\{q^T p\} \Leftrightarrow 0 \leq \min_{p\in\mathbb{R}^\kappa}\{q^T p : \exists W \succeq 0 : Qp = \mathcal{A}W\} \\ \Leftrightarrow 0 \leq \min_{p,W}\{q^T p : Qp - \mathcal{A}W = 0, W \succeq 0\} \\ \Leftrightarrow \{q = Q^T\lambda : \lambda \in \mathbb{R}^{2\kappa+1}, \mathcal{A}^*\lambda \succeq 0\}, \end{array}$$

where the concluding $\Leftrightarrow$ is due to semidefinite duality. Computing $\mathcal{A}^*\lambda$, we arrive at (5.3.9). □

Remark 5.1 *Note that* **C.2** *admits a straightforward modification where the spaces $\mathcal{F}_\ell$ are comprised of trigonometric polynomials $\sum_{i=0}^{\kappa}[p_i\cos(i\omega_\ell s) + q_i\sin(i\omega_\ell s)]$ rather than of algebraic polynomials $\sum_{i=0}^\kappa p_i s^i$. Here all we need is a tractable description of the convex hull of the curve*

$$\{[s;\cos(\omega_\ell s);\sin(\omega_\ell s);...;\cos(\kappa\omega_\ell s);\sin(\kappa\omega_\ell s)] : -1 \leq s \leq 1\}$$

which can be easily extracted from the semidefinite representation of the cone P_+.

Discussion. There are items to note on the results stated in **C**. The bad news is that *understood literally, these results have no direct consequences in our context — when $\mathcal{Z}$ is a box, decision rules* (5.3.7) *never outperform "genuine" affine decision rules with the same information base* (that is, the decision rules (5.3.7) with the spaces of affine functions on the axis in the role of $\mathcal{F}_\ell$).

The explanation is as follows. Consider, instead of (5.3.6), a more general problem, specifically, the uncertain problem

$$\mathcal{C} = \left\{\min_x \left\{c_\zeta^T x + d_\zeta : A_\zeta x - b_\zeta \in \mathbf{K}\right\} : \zeta \in \mathcal{Z}\right\} \quad [c_\zeta, d_\zeta, A_\zeta, b_\zeta : \text{ affine in } \zeta] \tag{5.3.11}$$

where $\mathbf{K}$ is a convex set. Assume that $\mathcal{Z}$ is a direct product of simplexes: $\mathcal{Z} = \Delta_1 \times ... \times \Delta_L$, where Δ_ℓ is a k_ℓ-dimensional simplex (the convex hull of $k_\ell + 1$ affinely independent points in $\mathbb{R}^{k_\ell}$). Assume we want to process the ARC of this problem restricted to the decision rules of the form

$$x_j = \xi_j + \sum_{\ell \in I_j} f_\ell^j(\zeta_\ell), \tag{5.3.12}$$

where ζ_ℓ is the projection of $\zeta \in \mathcal{Z}$ on Δ_ℓ, and the only restriction on the functions f_ℓ^j is that they belong to given families $\mathcal{F}_\ell$ of functions on $\mathbb{R}^{k_\ell}$. We still assume fixed recourse: the columns of A_ζ and the entries in c_ζ associated with adjustable, (i.e., with $I_j \neq \emptyset$) decision variables x_j are independent of ζ.

The above claim that "genuinely affine" decision rules are not inferior as compared to the rules (5.3.7) is nothing but the following simple observation:

Lemma 5.3 *Whenever certain $t \in \mathbb{R}$ is an achievable value of the objective in the ARC of* (5.3.11) *restricted to the decision rules* (5.3.12), *that is, there exist decision rules of the latter form such that*

$$\left.\begin{array}{l} \sum\limits_{j=1}^n \left[\xi_j + \sum\limits_{\ell \in I_j} f_\ell^j(\zeta_\ell)\right](c_\zeta)_j + d_\zeta \leq t \\ \sum\limits_{j=1}^n \left[\xi_j + \sum\limits_{\ell \in I_j} f_\ell^j(\zeta_\ell)\right] A_\zeta^j - b_\zeta \in \mathbf{K} \end{array}\right\} \quad \forall \zeta \begin{array}{l} \in [\zeta_1; ...; \zeta_L] \in \mathcal{Z} \\ = \Delta_1 \times ... \times \Delta_L, \end{array} \tag{5.3.13}$$

t is also an achievable value of the objective in the ARC of the uncertain problem restricted to affine decision rules with the same information base: there exist affine in ζ_ℓ functions $\phi_\ell^j(\zeta_\ell)$ such that (5.3.13) *remains valid with ϕ_ℓ^j in the role of f_ℓ^j.*

Proof is immediate: since every collection of $k_\ell + 1$ reals can be obtained as the collection of values of an affine function at the vertices of k_ℓ-dimensional simplex, we can find affine functions $\phi_\ell^j(\zeta_\ell)$ such that $\phi_\ell^j(\zeta_\ell) = f_\ell^j(\zeta_\ell)$ whenever ζ_ℓ is a vertex of the simplex Δ_ℓ. When plugging into the left hand sides of the constraints in (5.3.13) the functions $\phi_\ell^j(\zeta_\ell)$ instead of $f_\ell^j(\zeta_\ell)$, these left hand sides become affine functions of ζ (recall that we are in the case of fixed recourse). Due to this affinity and to the fact that $\mathcal{Z}$ is a convex compact set, in order for the resulting constraints to be valid for all $\zeta \in \mathcal{Z}$, it suffices for them to be valid at every one of the extreme points of $\mathcal{Z}$. The components $\zeta_1, ..., \zeta_L$ of such an extreme point ζ are vertices of $\Delta_1, ..., \Delta_L$, and therefore the validity of "ϕ constraints" at ζ is readily given by the validity of the "f constraints" at this point — by construction, at such a point the left hand sides of the "ϕ" and the "f" constraints coincide with each other. □

Does the bad news mean that our effort in **C.1–2** was just wasted? The good news is that this effort still can be utilized. Consider again the case where ζ_ℓ are scalars, assume that $\mathcal{Z}$ is not a box, in which case Lemma 5.3 is not applicable. Thus, we have hope that the ARC of (5.3.6) restricted to the decision rules (5.3.7) is indeed less conservative (has a strictly less optimal value) than the ARC restricted to the affine decision rules. What we need in order to process the former, "more promising," ARC, is a tractable description of the convex hull $\widehat{\mathcal{Z}}$ of the image $\mathcal{U}$ of $\mathcal{Z}$ under the mapping

$$\zeta \mapsto \widehat{\zeta}[\zeta] = \{\zeta_{\ell i}[\zeta_\ell]\}_{\substack{0 \leq i \leq m, \\ 1 \leq \ell \leq L}}$$

where $\zeta_{\ell 0} = \zeta_\ell$, $\zeta_{\ell i}[\zeta_\ell] = f_{i\ell}(\zeta_\ell)$, $1 \leq i \leq m$, and the functions $f_{i\ell} \in \mathcal{F}_\ell$, $i = 1, ..., m$, span $\mathcal{F}_\ell$. The difficulty is that *with $\mathcal{F}_\ell$ as those considered in* **C.1–2** (these families are "rich enough" for most of applications), *we, as a matter of fact, do not know how to get a tractable representation of $\widehat{\mathcal{Z}}$, unless $\mathcal{Z}$ is a box.* Thus, $\mathcal{Z}$ more complicated than a box seems to be too complex, and when $\mathcal{Z}$ is a box, we gain nothing from allowing for "complex" $\mathcal{F}_\ell$. Nevertheless, we can proceed as follows. Let us include $\mathcal{Z}$, (which is not a box), into a box $\mathcal{Z}^+$, and let us apply the outlined approach to $\mathcal{Z}^+$ in the role of $\mathcal{Z}$, that is, let us try to build a tractable description of the convex hull $\widehat{\mathcal{Z}}^+$ of the image $\mathcal{U}^+$ of $\mathcal{Z}^+$ under the mapping $\zeta \mapsto \widehat{\zeta}[\zeta]$. With luck, (e.g., in situations **C.1–2**), we will succeed, thus getting a tractable representation of $\widehat{\mathcal{Z}}^+$; the latter set is, of course, larger than the "true" set $\widehat{\mathcal{Z}}$ we want to describe. There is another "easy to describe" set that contains $\widehat{\mathcal{Z}}$, namely, the inverse image $\widehat{\mathcal{Z}}^0$ of $\mathcal{Z}$ under the natural projection $\Pi : \{\zeta_{\ell i}\}_{\substack{0\leq i\leq m,\\ 1\leq \ell\leq L}} \mapsto \{\zeta_{\ell 0}\}_{1\leq\ell\leq L}$ that recovers ζ from $\widehat{\zeta}[\zeta]$. And perhaps we are smart enough to find other easy to describe convex sets $\widehat{\mathcal{Z}}^1,...,\widehat{\mathcal{Z}}^k$ that contain $\widehat{\mathcal{Z}}$.

Assume, e.g., that $\mathcal{Z}$ is the Euclidean ball $\{\|\zeta\|_2 \leq r\}$, and let us take as $\mathcal{Z}^+$ the embedding box $\{\|\zeta\|_\infty \leq r\}$.

In the case of **C.1** we have for $i \geq 1$: $\zeta_{\ell i}[\zeta_\ell] = \max[\zeta_\ell, a_{\ell i}]$, whence $|\zeta_{\ell i}[\zeta_\ell]| \leq \max[|\zeta_\ell|, |a_{\ell i}|]$. It follows that when $\zeta \in \mathcal{Z}$, we have $\sum_\ell \zeta_{\ell i}^2[\zeta_\ell] \leq \sum_\ell \max[\zeta_\ell^2, a_{\ell i}^2] \leq \sum_\ell [\zeta_\ell^2 + a_{\ell i}^2] \leq r^2 + \sum_\ell a_{\ell i}^2$, and we can take as $\widehat{\mathcal{Z}}^p$, $p = 1, ..., m$, the elliptic cylinders $\{\{\zeta_{\ell i}\}_{\ell,i} : \sum_\ell \zeta_{\ell p}^2 \leq r^2 + \sum_\ell a_{\ell p}^2\}$. In the case of **C.2**, we have $\zeta_{\ell i}[\zeta_\ell] = \zeta_\ell^{i+1}$, $1 \leq i \leq \kappa - 1$, so that $\sum_\ell |\zeta_{\ell i}[\zeta_\ell]| \leq \max_{z\in\mathbb{R}^L}\{\sum_\ell |z_\ell|^{i+1} : \|z\|_2 \leq r\} = r^{i+1}$. Thus, we can take $\widehat{\mathcal{Z}}^p = \{\{\zeta_{\ell i}\}_{\ell,i} : \sum_\ell |\zeta_{\ell p}| \leq r^{p+1}\}$, $1 \leq p \leq \kappa - 1$.

Since all the easy to describe convex sets $\widehat{\mathcal{Z}}^+$, $\widehat{\mathcal{Z}}^0$,...,$\widehat{\mathcal{Z}}^k$ contain $\widehat{\mathcal{Z}}$, the same is true for the easy to describe convex set

$$\widetilde{\mathcal{Z}} = \widehat{\mathcal{Z}}^+ \cap \widehat{\mathcal{Z}}^0 \cap \widehat{\mathcal{Z}}^1 \cap ... \cap \widehat{\mathcal{Z}}^k,$$

so that the (tractable along with $\widetilde{\mathcal{Z}}$) semi-infinite LO problem

$$\min_{\substack{t,\\ \{X_j(\cdot)\in\mathcal{X}_j\}_{j=1}^n}} \left\{ t : \begin{array}{l} d_{\Pi(\widehat{\zeta})} + \sum\limits_{j=1}^n X_j(\widehat{\zeta})(c_{\Pi(\widehat{\zeta})})_j \leq t \\ \sum\limits_{j=1}^n X_j(\widehat{\zeta})A^j_{\Pi(\widehat{\zeta})} - b_{\Pi(\widehat{\zeta})} \geq 0 \end{array} \right\} \forall \widehat{\zeta} = \{\zeta_{\ell i}\} \in \widetilde{\mathcal{Z}} \right\} \tag{S}$$
$$\left[\Pi\left(\{\zeta_{\ell i}\}_{\substack{0\leq i\leq m,\\ 1\leq\ell\leq L}} \right) = \{\zeta_{\ell 0}\}_{1\leq\ell\leq L},\ \mathcal{X}_j = \{X_j(\widehat{\zeta}) = \xi_j + \sum_{\substack{\ell\in I_j,\\ 0\leq i\leq m}} \eta_{\ell i}\zeta_{\ell i}\} \right]$$

is a safe tractable approximation of the ARC of (5.3.6) restricted to decision rules (5.3.7). Note that this approximation is *at least* as flexible as the ARC of (5.3.6) restricted to genuine affine decision rules. Indeed, a rule $X(\cdot) = \{X_j(\cdot)\}_{j=1}^n$ of the latter type is "cut off" the family of all decision rules participating in (S) by the requirement "X_j depend solely on $\zeta_{\ell 0}$, $\ell \in I_j$," or, which is the same, by the requirement $\eta_{\ell i} = 0$ whenever $i > 0$. Since by construction the projection of $\widetilde{\mathcal{Z}}$ on the space of variables $\zeta_{\ell 0}$, $1 \leq \ell \leq L$, is exactly $\mathcal{Z}$, a pair $(t, X(\cdot))$ is feasible for (S) if and only if it is feasible for the AARC of (5.3.6), the information base being given by $I_1, ..., I_n$. The bottom line is, that when $\mathcal{Z}$ is not a box, the tractable problem (S), while still producing robust feasible decisions, is at least as flexible as the AARC. Whether this "at least as flexible" is or is not "more flexible," depends on the application in question, and since both (S) and AARC are tractable, it is easy to figure out what the true answer is.

Here is a toy example. Let $L = 2$, $n = 2$, and let (5.3.6) be the uncertain problem

$$\left\{ \min_x \left\{ x_2 : \begin{array}{l} x_1 \geq \zeta_1 \\ x_1 \geq -\zeta_1 \\ x_2 \geq x_1 + 3\zeta_1/5 + 4\zeta_2/5 \\ x_2 \geq x_1 - 3\zeta_1/5 - 4\zeta_2/5 \end{array} \right\}, \|\zeta\|_2 \leq 1 \right\},$$

with fully adjustable variable x_1 and non-adjustable variable x_2. Due to the extreme simplicity of our problem, we can immediately point out an optimal solution to the *unrestricted* ARC, namely,

$$X_1(\zeta) = |\zeta_1|,\ x_2 \equiv \mathrm{Opt}(\mathrm{ARC}) = \max_{\|\zeta\|_2 \le 1}[|\zeta_1| + |3\zeta_1 + 4\zeta_2|/5] = \frac{4\sqrt{5}}{5} \approx 1.7889.$$

Now let us compare Opt(ARC) with the optimal value Opt(AARC) of the AARC and with the optimal value Opt(RARC) of the restricted ARC where the decision rules are allowed to be affine in $[\zeta_\ell]_\pm$, $\ell = 1, 2$ (as always, $[a]_+ = \max[a, 0]$ and $[a]_- = \max[-a, 0]$). The situation fits **B**, so that we can process the RARC as it is. Noting that $a = [a]_+ - [a]_-$, the decision rules that are affine in $[\zeta_\ell]_\pm$, $\ell = 1, 2$, are exactly the same as the decision rules (5.3.7), where $\mathcal{F}_\ell$, $\ell = 1, 2$, are the spaces of piecewise linear functions on the axis with the only breakpoint 0. We see that *up to the fact that $\mathcal{Z}$ is a circle rather than a square*, the situation fits **C.1** as well, and we can process RARC via its safe tractable approximation (S). Let us look what are the optimal values yielded by these 3 schemes.

• The AARC of our toy problem is

$$\mathrm{Opt}(\mathrm{AARC}) = \min_{x_2,\xi,\eta}\left\{x_2 : \begin{array}{ll} \overbrace{\xi + \eta^T\zeta}^{X_1(\zeta)} \ge |\zeta_1| & (a) \\ x_2 \ge X_1(\zeta) + |3\zeta_1 + 4\zeta_2|/5 & (b) \end{array} \quad \forall(\zeta : \|\zeta\|_2 \le 1)\right\}$$

This problem can be immediately solved. Indeed, (a) should be valid for $\zeta = \zeta^1 \equiv [1; 0]$ and for $\zeta = \zeta^2 \equiv -\zeta^1$, meaning that $X_1(\pm\zeta^1) \ge 1$, whence $\xi \ge 1$. Further, (b) should be valid for $\zeta = \zeta^3 \equiv [3; 4]/5$ and for $\zeta = \zeta^4 \equiv -\zeta^3$, meaning that $x_2 \ge X_1(\pm\zeta^3) + 1$, whence $x_2 \ge \xi + 1 \ge 2$. We see that the optimal value is ≥ 2, and this bound is achievable (we can take $X_1(\cdot) \equiv 1$ and $x_2 = 2$). As a byproduct, in our toy problem the AARC is as conservative as the RC.

• The RARC of our problem as given by **B** is

$$\mathrm{Opt}(\mathrm{RARC}) = \min_{x_2,\xi,\eta,\eta_\pm}\left\{x_2 : \begin{array}{l} \overbrace{\xi + \eta^T\zeta + \eta_+^T\zeta^+ + \eta_-^T\zeta^-}^{X_1(\widehat{\zeta})} \ge |\zeta_1| \\ x_2 \ge X_1(\widehat{\zeta}) + |3\zeta_1 + 4\zeta_2|/5 \end{array} \quad \forall(\widehat{\zeta} = [\underbrace{\zeta_1; \zeta_2}_{\zeta}; \underbrace{\zeta_1^+; \zeta_2^+}_{\zeta^+}; \underbrace{\zeta_1^-; \zeta_2^-}_{\zeta^-}] \in \widehat{\mathcal{Z}})\right\},$$

$$\widehat{\mathcal{Z}} = \left\{\widehat{\zeta} : \zeta = \zeta^+ - \zeta^-, \zeta^\pm \ge 0, \|\zeta^+ + \zeta^-\|_2 \le 1\right\}.$$

We can say in advance what are the optimal value and the optimal solution to the RARC — they should be the same as those of the ARC, since the latter, as a matter of fact, admits optimal decision rules that are affine in $|\zeta_1|$, and thus in $[\zeta_\ell]_\pm$. Nevertheless, we have carried out numerical optimization which yielded another optimal solution to the RARC (and thus - to ARC):

$$\begin{array}{l} \mathrm{Opt}(\mathrm{RARC}) = x_2 = 1.7889, \\ \xi = 1.0625, \eta = [0; 0], \eta_+ = \eta_- = [0.0498; -0.4754], \end{array}$$

which corresponds to $X_1(\zeta) = 1.0625 + 0.0498|\zeta_1| - 0.4754|\zeta_2|$.

• The safe tractable approximation of the RARC looks as follows. The mapping $\zeta \mapsto \widehat{\zeta}[\zeta]$ in our case is

$$[\zeta_1; \zeta_2] \mapsto [\zeta_{1,0} = \zeta_1; \zeta_{1,1} = \max[\zeta_1, 0]; \zeta_{2,0} = \zeta_2; \zeta_{2,1} = \max[\zeta_2, 0]],$$

the tractable description of $\widehat{\mathcal{Z}}^+$ as given by **C.1** is

$$\widehat{\mathcal{Z}}^+ = \left\{ \{\zeta_{\ell i}\}_{\substack{i=0,1\\ \ell=1,2}} : \begin{array}{c} -1 \le \zeta_{\ell 0} \le 1 \\ 0 \le \frac{\zeta_{\ell 1}-\zeta_{\ell 0}}{1} \le \frac{1-\zeta_{\ell 1}}{1} \le 1 \end{array} \right\}, \ell = 1,2\Big\}$$

and the sets $\widehat{\mathcal{Z}}^0$, $\widehat{\mathcal{Z}}^1$ are given by

$$\widehat{\mathcal{Z}}^i = \left\{ \{\zeta_{\ell i}\}_{\substack{i=0,1\\ \ell=1,2}} : \zeta_{1i}^2 + \zeta_{2i}^2 \le 1 \right\}, \; i = 0,1.$$

Consequently, (S) becomes the semi-infinite LO problem

$$\mathrm{Opt}(S) \;=\; \min_{x_2,\xi,\{\eta_{\ell i}\}} \left\{ x_2 : \begin{array}{l} X_1(\widehat{\zeta}) \equiv \xi + \sum_{\substack{\ell=1,2\\ i=0,1}} \eta_{\ell i}\zeta_{\ell i} \ge \zeta_{1,0} \\ X_1(\widehat{\zeta}) \equiv \xi + \sum_{\substack{\ell=1,2\\ i=0,1}} \eta_{\ell i}\zeta_{\ell i} \ge -\zeta_{1,0} \\ x_2 \ge \xi + \sum_{\substack{\ell=1,2\\ i=0,1}} \eta_{\ell i}\zeta_{\ell i} + [3\zeta_{1,0} + 4\zeta_{2,0}]/5 \\ x_2 \ge \xi + \sum_{\substack{\ell=1,2\\ i=0,1}} \eta_{\ell i}\zeta_{\ell i} - [3\zeta_{1,0} + 4\zeta_{2,0}]/5 \end{array} \right. \\ \left. \forall \widehat{\zeta} = \{\zeta_{\ell i}\} : \begin{array}{l} -1 \le \zeta_{\ell 0} \le 1, \ell = 1,2 \\ 0 \le \zeta_{\ell 1} - \zeta_{\ell 0} \le 1 - \zeta_{\ell 1} \le 1, \ell = 1,2 \\ \zeta_{1i}^2 + \zeta_{2i}^2 \le 1, \; i = 0,1 \end{array} \right\}.$$

Computation results in

$$\begin{array}{l} \mathrm{Opt}(S) = x_2 = \frac{25+\sqrt{8209}}{60} \approx 1.9267, \\ X_1(\zeta) = \frac{5}{12} - \frac{3}{5}\zeta_{1.0}[\zeta_1] + \frac{6}{5}\zeta_{1,1}[\zeta_1] + \frac{7}{60}\zeta_{2,0}[\zeta_2] = \frac{5}{12} + \frac{3}{5}|\zeta_1| + \frac{7}{60}\zeta_2. \end{array}$$

As it could be expected, we get $2 = \mathrm{Opt(AARC)} > 1.9267 = \mathrm{Opt}(S) > 1.7889 = \mathrm{Opt(RARC)} = \mathrm{Opt(ARC)}$. Note that in order to get $\mathrm{Opt}(S) < \mathrm{Opt(AARC)}$, taking into account $\widehat{\mathcal{Z}}^1$ is a must: in the case of **C.1**, whatever be $\mathcal{Z}$ and a box $\mathcal{Z}^+ \supset \mathcal{Z}$, with $\widetilde{\mathcal{Z}} = \widehat{\mathcal{Z}}^+ \cap \widehat{\mathcal{Z}}^0$ we gain nothing as compared to the genuine affine decision rules.

D. *Quadratic decision rules, ellipsoidal uncertainty set.* In this case,

$$\widehat{\zeta}[\zeta] = \left[\begin{array}{c|c} & \zeta^T \\ \hline \zeta & \zeta\zeta^T \end{array}\right]$$

is comprised of the entries of ζ and their pairwise products (so that the associated decision rules (5.3.5) are quadratic in ζ), and $\mathcal{Z}$ is the ellipsoid $\{\zeta \in \mathbb{R}^L : \|Q\zeta\|_2 \le 1\}$, where Q has a trivial kernel. The convex hull of the image of $\mathcal{Z}$ under the quadratic mapping $\zeta \to \widehat{\zeta}[\zeta]$ is easy to describe:

Lemma 5.4 *In the above notation, the set* $\widehat{\mathcal{Z}} = \mathrm{Conv}\{\widehat{\zeta}[\zeta] : \|Q\zeta\|_2 \le 1\}$ *is a convex compact set given by the semidefinite representation as follows:*

$$\widehat{\mathcal{Z}} = \left\{ \widehat{\zeta} = \left[\begin{array}{c|c} & v^T \\ \hline v & W \end{array}\right] \in \mathbf{S}^{L+1} : \widehat{\zeta} + \left[\begin{array}{c|c} 1 & \\ \hline & \end{array}\right] \succeq 0, \mathrm{Tr}(QWQ^T) \le 1 \right\}.$$

Proof. It is immediately seen that it suffices to prove the statement when $Q = I$, which we assume from now on. Besides this, when we add to the mapping $\widehat{\zeta}[\zeta]$ the constant matrix $\left[\begin{array}{c|c} 1 & \\ \hline & \end{array}\right]$, the convex hull of the image of $\mathcal{Z}$ is translated by the same matrix. It follows that all we need is to prove that the convex hull $\mathcal{Q}$ of the image of the unit Euclidean ball under the mapping $\zeta \mapsto \widetilde{\zeta}[\zeta] = \left[\begin{array}{c|c} 1 & \zeta^T \\ \hline \zeta & \zeta\zeta^T \end{array}\right]$ can be represented as

$$\mathcal{Q} = \left\{ \widehat{\zeta} = \left[\begin{array}{c|c} 1 & v^T \\ \hline v & W \end{array}\right] \in \mathbf{S}^{L+1} : \widehat{\zeta} \succeq 0, \mathrm{Tr}(W) \le 1 \right\}. \quad (5.3.14)$$

Denoting the right hand side in (5.3.14) by $\widehat{\mathcal{Q}}$, both $\mathcal{Q}$ and $\widehat{\mathcal{Q}}$ are nonempty convex compact sets. Therefore they coincide if and only if their support functions are identical.[1] We are in the situation where $\mathcal{Q}$ is the convex hull of the set $\left\{\left[\begin{array}{c|c} 1 & \zeta^T \\ \hline \zeta & \zeta\zeta^T \end{array}\right] : \zeta^T\zeta \leq 1\right\}$, so that the support function of $\mathcal{Q}$ is

$$\phi(P) = \max_Z \left\{ \mathrm{Tr}(PZ) : Z = \left[\begin{array}{c|c} 1 & \zeta^T \\ \hline \zeta & \zeta\zeta^T \end{array}\right] : \zeta^T\zeta \leq 1 \right\} \qquad \left[P = \left[\begin{array}{c|c} p & q^T \\ \hline q & R \end{array}\right] \in \mathbf{S}^{L+1}\right].$$

We have

$$\begin{array}{l}
\phi(P) = \max_Z \left\{ \mathrm{Tr}(PZ) : Z = \left[\begin{array}{c|c} 1 & \zeta^T \\ \hline \zeta & \zeta\zeta^T \end{array}\right] \text{ with } \zeta^T\zeta \leq 1 \right\} \\
= \max_\zeta \left\{ \zeta^T R\zeta + 2q^T\zeta + p : \zeta^T\zeta \leq 1 \right\} \\
= \min_\tau \left\{ \tau : \tau \geq \zeta^T R\zeta + 2q^T\zeta + p \,\forall(\zeta : \zeta^T\zeta \leq 1) \right\} \\
= \min_\tau \left\{ \tau : (\tau - p)t^2 - \zeta^T R\zeta - 2tq^T\zeta \geq 0 \,\forall((\zeta, t) : \zeta^T\zeta \leq t^2) \right\} \\
= \min_\tau \left\{ \tau : \exists \lambda \geq 0 : (\tau - p)t^2 - \zeta^T R\zeta - 2tq^T\zeta - \lambda(t^2 - \zeta^T\zeta) \geq 0 \,\forall(\zeta, t) \right\} \ [\mathcal{S}\text{-Lemma}] \\
= \min_{\tau,\lambda} \left\{ \tau : \left[\begin{array}{c|c} \tau - p - \lambda & -q^T \\ \hline -q & \lambda I - R \end{array}\right] \succeq 0, \lambda \geq 0 \right\} \\
= \max_{u,v,W,r} \left\{ up + 2v^T q + \mathrm{Tr}(RW) : \mathrm{Tr}\left(\left[\begin{array}{c|c} \tau - \lambda & \\ \hline & \lambda I \end{array}\right]\left[\begin{array}{c|c} u & v^T \\ \hline v & W \end{array}\right]\right) + r\lambda \right. \\
\qquad \left. \equiv \tau \forall(\tau, \lambda), \left[\begin{array}{c|c} u & v^T \\ \hline v & W \end{array}\right] \succeq 0, r \geq 0 \right\} \ [\text{semidefinite duality}] \\
= \max_{v,W} \left\{ p + 2v^T q + \mathrm{Tr}(RW) : \left[\begin{array}{c|c} 1 & v^T \\ \hline v & W \end{array}\right] \succeq 0, \mathrm{Tr}(W) \leq 1 \right\} \\
= \max_{v,W} \left\{ \mathrm{Tr}\left(P\left[\begin{array}{c|c} 1 & v^T \\ \hline v & W \end{array}\right]\right) : \left[\begin{array}{c|c} 1 & v^T \\ \hline v & W \end{array}\right] \in \widehat{\mathcal{Q}} \right\}.
\end{array}$$

Thus, the support function of $\mathcal{Q}$ indeed is identical to the one of $\widehat{\mathcal{Q}}$. □

Corollary 5.1 *Consider a fixed recourse uncertain LO problem* (5.3.6) *with an ellipsoid as an uncertainty set, where the adjustable decision variables are allowed to be quadratic functions of prescribed portions $P_j\zeta$ of the data. The associated ARC of the problem is computationally tractable and is given by an explicit semidefinite program of the sizes polynomial in those of instances and in the dimension L of the data vector.*

E. *Quadratic decision rules and the intersection of concentric ellipsoids as the uncertainty set.* Here the uncertainty set $\mathcal{Z}$ is $\cap$-ellipsoidal:

$$\begin{array}{c} \mathcal{Z} = \mathcal{Z}_\rho \equiv \{\zeta \in \mathbb{R}^L : \zeta^T Q_j \zeta \leq \rho^2, 1 \leq j \leq J\} \\ \left[Q_j \succeq 0, \sum_j Q_j \succ 0\right] \end{array} \tag{5.3.15}$$

(cf. section 3.3.2), where $\rho > 0$ is an uncertainty level, and, as above, $\widehat{\zeta}[\zeta] = \left[\begin{array}{c|c} & \zeta^T \\ \hline \zeta & \zeta\zeta^T \end{array}\right]$, so that our intention is to process the ARC of an uncertain problem corresponding to quadratic decision rules. As above, all we need is to get a tractable representation of the convex hull of

[1] The support function of a nonempty convex set $X \subset \mathbb{R}^n$ is the function $f(\xi) = \sup_{x \in X} \xi^T x : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$. The fact that two closed nonempty convex sets in $\mathbb{R}^n$ are identical, if and only if their support functions are so, is readily given by the Separation Theorem.

the image of $\mathcal{Z}_\rho$ under the nonlinear mapping $\zeta \mapsto \widehat{\zeta}[\zeta]$. This is essentially the same as to find a similar representation of the convex hull $\widehat{\mathcal{Z}}_\rho$ of the image of $\mathcal{Z}_\rho$ under the nonlinear mapping

$$\zeta \mapsto \widehat{\zeta}_\rho[\zeta] = \left[\begin{array}{c|c} & \zeta^T \\ \hline \zeta & \frac{1}{\rho}\zeta\zeta^T \end{array}\right];$$

indeed, both convex hulls in question can be obtained from each other by simple linear transformations. The advantage of our normalization is that now $\mathcal{Z}_\rho = \rho\mathcal{Z}_1$ and $\widehat{\mathcal{Z}}_\rho = \rho\widehat{\mathcal{Z}}_1$, as it should be for respectable perturbation sets.

While the set $\widehat{\mathcal{Z}}_\rho$ is, in general, computationally intractable, we are about to demonstrate that this set admits a tight tractable approximation, and that the latter induces a tight tractable approximation of the "quadratically adjustable" RC of the Linear Optimization problem in question. The main ingredient we need is as follows:

Lemma 5.5 *Consider the semidefinite representable set*

$$\mathcal{W}_\rho = \rho\mathcal{W}_1,\ \mathcal{W}_1 = \left\{\widehat{\zeta} = \left[\begin{array}{c|c} & v^T \\ \hline v & W \end{array}\right] : \left[\begin{array}{c|c} 1 & v^T \\ \hline v & W \end{array}\right] \succeq 0, \mathrm{Tr}(WQ_j) \le 1,\ 1 \le j \le J\right\}. \tag{5.3.16}$$

Then

$$\forall \rho > 0 : \widehat{\mathcal{Z}}_\rho \subset \mathcal{W}_\rho \subset \widehat{\mathcal{Z}}_{\vartheta\rho}, \tag{5.3.17}$$

where $\vartheta = O(1)\ln(J+1)$ and J is the number of ellipsoids in the description of $\mathcal{Z}_\rho$.

Proof. Since both $\widehat{\mathcal{Z}}_\rho$ and $\widehat{\mathcal{W}}_\rho$ are nonempty convex compact sets containing the origin and belonging to the subspace $\mathbf{S}_0^{L+1}$ of $\mathbf{S}^{L+1}$ comprised of matrices with the first diagonal entry being zero, to prove (5.3.17) is the same as to verify that the corresponding support functions

$$\phi_{\mathcal{W}_\rho}(P) = \max_{\widehat{\zeta}\in\mathcal{W}_\rho} \mathrm{Tr}(P\widehat{\zeta}),\ \phi_{\widehat{\mathcal{Z}}_\rho}(P) = \max_{\widehat{\zeta}\in\widehat{\mathcal{Z}}_\rho} \mathrm{Tr}(P\widehat{\zeta}),$$

considered as functions of $P \in \mathbf{S}_0^{L+1}$, satisfy the relation

$$\phi_{\widehat{\mathcal{Z}}_\rho}(\cdot) \le \phi_{\mathcal{W}_\rho}(\cdot) \le \phi_{\widehat{\mathcal{Z}}_{\theta\rho}}(\cdot).$$

Taking into account that $\widehat{\mathcal{Z}}_s = s\widehat{\mathcal{Z}}_1$, $s > 0$, this task reduces to verifying that

$$\phi_{\widehat{\mathcal{Z}}_\rho}(\cdot) \le \phi_{\mathcal{W}_\rho}(\cdot) \le \vartheta\phi_{\widehat{\mathcal{Z}}_\rho}(\cdot).$$

Thus, all we should prove is that whenever $P = \left[\begin{array}{c|c} & p^T \\ \hline p & R \end{array}\right] \in \mathbf{S}_0^{L+1}$, one has

$$\max_{\widehat{\zeta}\in\widehat{\mathcal{Z}}_\rho} \mathrm{Tr}(P\widehat{\zeta}) \le \max_{\widehat{\zeta}\in\mathcal{W}_\rho} \mathrm{Tr}(P\widehat{\zeta}) \le \vartheta \max_{\widehat{\zeta}\in\widehat{\mathcal{Z}}_\rho} \mathrm{Tr}(P\widehat{\zeta}).$$

Recalling the origin of $\widehat{\mathcal{Z}}_\rho$, the latter relation reads

$$\begin{array}{c} \forall P = \left[\begin{array}{c|c} & p^T \\ \hline p & R \end{array}\right] : \mathrm{Opt}_P(\rho) \equiv \max\limits_{\zeta}\left\{2p^T\zeta + \frac{1}{\rho}\zeta^T R\zeta : \zeta^T Q_j \zeta \le \rho^2, 1 \le j \le J\right\} \\ \le \mathrm{SDP}_P(\rho) \equiv \max\limits_{\widehat{\zeta}\in\mathcal{W}_\rho} \mathrm{Tr}(P\widehat{\zeta}) \le \vartheta \mathrm{Opt}_P(\rho) \equiv \mathrm{Opt}_P(\vartheta\rho). \end{array} \tag{5.3.18}$$

Observe that the three quantities in the latter relation are of the same homogeneity degree w.r.t. $\rho > 0$, so that it suffices to verify this relation when $\rho = 1$, which we assume from now on.

We are about to derive (5.3.18) from the Approximate $\mathcal{S}$-Lemma (Theorem A.8). To this end, let us specify the entities participating in the latter statement as follows:

- $x = [t; \zeta] \in \mathbb{R}^1_t \times \mathbb{R}^L_\zeta$;
- $A = P$, that is, $x^T A x = 2tp^T\zeta + \zeta^T R\zeta$;
- $B = \left[\begin{array}{c|c} 1 & \\ \hline & \end{array}\right]$, that is, $x^T B x = t^2$;
- $B_j = \left[\begin{array}{c|c} & \\ \hline & Q_j \end{array}\right]$, $1 \le j \le J$, that is, $x^T B_j x = \zeta^T Q_j \zeta$;
- $\rho = 1$.

With this setup, the quantity $\mathrm{Opt}(\rho)$ from (A.4.12) becomes nothing but $\mathrm{Opt}_P(1)$, while the quantity $\mathrm{SDP}(\rho)$ from (A.4.13) is

$$\begin{array}{l}
\mathrm{SDP}(1) = \max\limits_X \{\mathrm{Tr}(AX) : \mathrm{Tr}(BX) \le 1, \mathrm{Tr}(B_j X) \le 1,\, 1 \le j \le J, X \succeq 0\}\\
= \max\limits_X \left\{ 2p^T v + \mathrm{Tr}(RW) : \begin{array}{l} u \le 1 \\ \mathrm{Tr}(WQ_j) \le 1,\, 1 \le j \le J \\ X = \left[\begin{array}{c|c} u & v^T \\ \hline v & W \end{array}\right] \succeq 0 \end{array} \right\}\\
= \max\limits_{v,W} \left\{ 2p^T v + \mathrm{Tr}(RW) : \begin{array}{l} \mathrm{Tr}(WQ_j) \le 1,\, 1 \le j \le J \\ \left[\begin{array}{c|c} 1 & v^T \\ \hline v & W \end{array}\right] \succeq 0 \end{array} \right\}\\
= \max\limits_{\widehat{\zeta}} \left\{ \mathrm{Tr}(P\widehat{\zeta}) : \widehat{\zeta} = \left[\begin{array}{c|c} & v^T \\ \hline v & W \end{array}\right] : \begin{array}{l} \left[\begin{array}{c|c} 1 & v^T \\ \hline v & W \end{array}\right] \succeq 0 \\ \mathrm{Tr}(WQ_j) \le 1,\, 1 \le j \le J \end{array} \right\}\\
= \mathrm{SDP}_P(1).
\end{array}$$

With these observations, the conclusion (A.4.15) of the Approximate $\mathcal{S}$-Lemma reads

$$\mathrm{Opt}_P(1) \le \mathrm{SDP}_P(1) \le \mathrm{Opt}(\Omega(J)),\ \Omega(J) = 9.19\sqrt{\ln(J+1)} \tag{5.3.19}$$

where for $\Omega \ge 1$

$$\begin{array}{l}
\mathrm{Opt}(\Omega) = \max\limits_x \left\{x^T A x : x^T B x \le 1, x^T B_j x \le \Omega^2\right\}\\
= \max\limits_{t,\zeta} \left\{2tp^T\zeta + \zeta^T R\zeta : t^2 \le 1, \zeta^T Q_j \zeta \le \Omega^2,\, 1 \le j \le J\right\}\\
= \max\limits_{\zeta} \left\{2p^T\zeta + \zeta^T R\zeta : \zeta^T Q_j \zeta \le \Omega^2,\, 1 \le j \le J\right\}\\
= \max\limits_{\eta = \Omega^{-1}\zeta} \left\{\Omega(2p^T\eta) + \Omega^2 \eta^T R\eta : \eta^T Q_j \eta \le 1,\, 1 \le j \le J\right\}\\
\le \Omega^2 \max\limits_{\eta} \left\{2p^T\eta + \eta^T R\eta : \eta^T Q_j \eta \le 1,\, 1 \le j \le J\right\}\\
= \Omega^2 \mathrm{Opt}_P(1).
\end{array}$$

Setting $\vartheta = \Omega^2(J)$, we see that (5.3.19) implies (5.3.18). □

Corollary 5.2 *Consider a fixed recourse uncertain LO problem* (5.3.6) *with* $\cap$*-ellipsoidal uncertainty set* $\mathcal{Z}_\rho$ *(see* (5.3.15)*) where one seeks robust optimal quadratic decision rules:*

$$x_j = p_j + q_j^T \widehat{P}_j\left(\widehat{\zeta}_\rho[\zeta]\right)$$
$$\left[\begin{array}{l}
\bullet\ \ \widehat{\zeta}_\rho[\zeta] = \left[\begin{array}{c|c} & \zeta^T \\ \hline \zeta & \frac{1}{\rho}\zeta\zeta^T \end{array}\right] \\
\bullet\ \ \widehat{P}_j : \ \textit{linear mappings from } \mathbf{S}^{L+1} \textit{ to } \mathbb{R}^{m_j} \\
\bullet\ \ p_j \in \mathbb{R}, q_j \in \mathbb{R}^{m_j} : \ \textit{parameters to be specified}
\end{array}\right]. \tag{5.3.20}$$

The associated Adjustable Robust Counterpart of the problem admits a safe tractable approximation that is tight within the factor ϑ *given by Lemma 5.5.*

Here is how the safe approximation of the Robust Counterpart mentioned in Corollary 5.2 can be built:

1. We write down the optimization problem

$$\min_{t,x}\left\{t:\begin{array}{l} a_{0\zeta}^T[t;x]+b_{0\zeta}\equiv t-c_\zeta^T x-d_\zeta\geq 0\\ a_{i\zeta}^T[t;x]+b_{i,\zeta}\equiv A_{i\zeta}^T x-b_{i\zeta}\geq 0,\ i=1,...,m\end{array}\right\} \tag{P}$$

   where $A_{i\zeta}^T$ is i-th row in A_ζ and $b_{i\zeta}$ is i-th entry in b_ζ;

2. We plug into the $m+1$ constraints of (P), instead of the original decision variables x_j, the expressions $p_j+q_j^T\widehat{P}_j\left(\widehat{\zeta}_\rho[\zeta]\right)$, thus arriving at the optimization problem of the form

$$\min_{[t;y]}\left\{t:\alpha_{i\widehat{\zeta}}^T[t;y]+\beta_{i\widehat{\zeta}}\geq 0,\ 0\leq i\leq m\right\}, \tag{P'}$$

   where y is the collection of coefficients p_j,q_j of the quadratic decision rules, $\widehat{\zeta}$ is our new uncertain data — a matrix from $\mathbf{S}_0^{L+1}$ (see p. 233), and $\alpha_{i\widehat{\zeta}}$, $\beta_{i\widehat{\zeta}}$ are affine in $\widehat{\zeta}$, the affinity being ensured by the assumption of fixed recourse. The "true" quadratically adjustable RC of the problem of interest is the semi-infinite problem

$$\min_{[t;y]}\left\{t:\forall\widehat{\zeta}\in\widehat{\mathcal{Z}}_\rho:\alpha_{i\widehat{\zeta}}^T[t;y]+\beta_{i\widehat{\zeta}}\geq 0,\ 0\leq i\leq m\right\} \tag{R}$$

   obtained from (P') by requiring the constraints to remain valid for all $\widehat{\zeta}\in\widehat{\mathcal{Z}}_\rho$, the latter set being the convex hull of the image of $\mathcal{Z}_\rho$ under the mapping $\zeta\mapsto\widehat{\zeta}_\rho[\zeta]$. The semi-infinite problem (R) in general is intractable, and we replace it with its safe tractable approximation

$$\min_{[t;y]}\left\{t:\forall\widehat{\zeta}\in\mathcal{W}_\rho:\alpha_{i\widehat{\zeta}}^T[t;y]+\beta_{i\widehat{\zeta}}\geq 0,\ 0\leq i\leq m\right\}, \tag{R'}$$

   where $\mathcal{W}_\rho$ is the semidefinite representable convex compact set defined in Lemma 5.5. By Theorem 1.1, (R') is tractable and can be straightforwardly converted into a semidefinite program of sizes polynomial in $n=\dim x$, m and $L=\dim\zeta$. Here is the conversion: recalling the structure of $\widehat{\zeta}$ and setting $z=[t;x]$, we can rewrite the body of i-th constraint in (R') as

$$\alpha_{i\widehat{\zeta}}^T z+\beta_{i\widehat{\zeta}}\equiv a_i[z]+\mathrm{Tr}(\underbrace{\left[\begin{array}{c|c} & v^T\\ \hline v & W\end{array}\right]}_{\widehat{\zeta}}\left[\begin{array}{c|c} & p_i^T[z]\\ \hline p_i[z] & P_i[z]\end{array}\right]),$$

   where $a_i[z]$, $p_i[z]$ and $P_i[z]=P_i^T[z]$ are affine in z. Therefore, invoking the definition of $\mathcal{W}_\rho=\rho\mathcal{W}_1$ (see Lemma 5.5), the RC of the i-th semi-infinite constraint in (R') is the first predicate in the following chain of equivalences:

$$\min_{v,W}\left\{\begin{array}{l}a_i[z]+2\rho v^T p_i[z]+\rho\mathrm{Tr}(WP_i[z]):\\ \qquad\left[\begin{array}{c|c}1 & v^T\\ \hline v & W\end{array}\right]\succeq 0,\mathrm{Tr}(WQ_j)\leq 1,\ 1\leq j\leq J\end{array}\right\}\geq 0 \tag{a_i}$$

$$\Updownarrow$$

$$\exists\lambda^i=[\lambda_1^i;...;\lambda_J^i]:\left\{\begin{array}{l}\lambda^i\geq 0\\ \left[\begin{array}{c|c}a_i[z]-\sum_j\lambda_j^i & \rho p_i^T[z]\\ \hline \rho p_i[z] & \rho P_i[z]+\sum_j\lambda_j^i Q_j\end{array}\right]\succeq 0\end{array}\right. \tag{b_i}$$

where $\Updownarrow$ is given by Semidefinite Duality. Consequently, we can reformulate (R') equivalently as the semidefinite program

$$\min_{\substack{z=[t;y]\\ \{\lambda_j^i\}}} \left\{ t : \begin{array}{l} \left[\begin{array}{c|c} a_i[z] - \sum_j \lambda_j^i & \rho p_i^T[z] \\ \hline \rho p_i[z] & \rho P_i[z] + \sum_j \lambda_j^i Q_j \end{array}\right] \succeq 0 \\ \lambda_j^i \geq 0,\, 0 \leq i \leq m, 1 \leq j \leq J \end{array} \right\}.$$

The latter SDP is a ϑ-tight safe tractable approximation of the quadratically adjustable RC with ϑ given by Lemma 5.5.

5.3.3 The AARC of Uncertain Linear Optimization Problem Without Fixed Recourse

We have seen that the AARC of an uncertain LO problem

$$\begin{array}{c} \mathcal{C} = \left\{ \min_x \left\{ c_\zeta^T x + d_\zeta : A_\zeta x \geq b_\zeta \right\} : \zeta \in \mathcal{Z} \right\} \\ [c_\zeta, d_\zeta, A_\zeta, b_\zeta : \text{ affine in } \zeta] \end{array} \tag{5.3.21}$$

with computationally tractable convex compact uncertainty set $\mathcal{Z}$ *and with fixed recourse* is computationally tractable. What happens when the assumption of fixed recourse is removed? The answer is that in general the AARC can become intractable (see [14]). However, we are about to demonstrate that for an ellipsoidal uncertainty set $\mathcal{Z} = \mathcal{Z}_\rho = \{\zeta : \|Q\zeta\|_2 \leq \rho\}$, $\mathrm{Ker} Q = \{0\}$, the AARC is computationally tractable, and for the $\cap$-ellipsoidal uncertainty set $\mathcal{Z} = \mathcal{Z}_\rho$ given by (5.3.15), the AARC admits a tight safe tractable approximation. Indeed, for affine decision rules

$$x_j = X_j(P_j\zeta) \equiv p_j + q_j^T P_j \zeta$$

the AARC of (5.3.21) is the semi-infinite problem of the form

$$\min_{z=[t;y]} \left\{ t : \forall \zeta \in \mathcal{Z}_\rho : a_i[z] + 2b_i^T[z]\zeta + \zeta^T C_i[z]\zeta \leq 0,\, 0 \leq i \leq m \right\}, \tag{5.3.22}$$

where $y = \{p_j, q_j\}_{j=1}^n$ and $a_i[z], b_i[z], C_i[z]$ are real/vector/symmetric matrix affinely depending on $z = [t; y]$. Consider the case of $\cap$-ellipsoidal uncertainty:

$$\mathcal{Z}_\rho = \{\zeta : \zeta^T Q_j \zeta \leq \rho^2,\, 1 \leq j \leq J\} \qquad [Q_j \succeq 0, \textstyle\sum_j Q_j \succ 0].$$

For a fixed i, $0 \leq i \leq m$, let us set $A_{i,z} = \left[\begin{array}{c|c} & b_i^T[z] \\ \hline b_i[z] & C_i[z] \end{array}\right]$, $B = \left[\begin{array}{c|c} 1 & \\ \hline & \end{array}\right]$, $B_j = \left[\begin{array}{c|c} & \\ \hline & Q_j \end{array}\right]$, $1 \leq j \leq J$, and observe that clearly

$$\begin{array}{l} \mathrm{Opt}_{i,z}(\rho) := \max\limits_{\eta=[\tau;\zeta]} \left\{ \eta^T A_{i,z}\eta \equiv 2\tau b_i^T[z]\zeta : \eta^T B\eta \equiv \tau^2 \leq 1, \eta^T B_j \eta \equiv \zeta^T Q_j \zeta \leq \rho^2, 1 \leq j \leq J \right\} \\ \leq \mathrm{SDP}_{i,z} := \min_{\lambda \geq 0} \left\{ \lambda_0 + \rho^2 \sum_{j=1}^J \lambda_j : \lambda_0 B + \sum_j \lambda_j B_j \succeq A_{i,z} \right\}, \end{array}$$

so that the explicit system of LMIs

$$\lambda_0 B + \sum_{j=1}^J \lambda_j B_j \succeq A_{i,z}, \lambda_0 + \rho^2 \sum_j \lambda_j \leq -a_i[z], \lambda \geq 0 \tag{5.3.23}$$

in variables z, λ is a safe tractable approximation of the i-th semi-infinite constraint

$$a_i[z] + 2b_i[z]\zeta + \zeta^T C[z]\zeta \leq 0 \ \forall \zeta \in \mathcal{Z}_\rho \tag{5.3.24}$$

appearing in (5.3.22). Let us prove that this approximation is tight within the factor ϑ which is 1 when $J = 1$ and is $9.19\sqrt{\ln(J)}$ otherwise. All we need to prove is that if z cannot be extended to a feasible solution of (5.3.23), z is infeasible for the semi-infinite constraint in question at the uncertainty level $\vartheta\rho$, or, which is clearly the same, that $\mathrm{Opt}_{i,z}(\vartheta\rho) > -a_i[z]$. When z cannot be extended to a feasible solution to (5.3.23), we have $\mathrm{SDP}_{i,z} > -a_i[z]$. Invoking Approximate $\mathcal{S}$-Lemma (Theorem A.8) for the data $A = A_{i,z}$, B, $\{B_j\}$, there exists $\bar\eta = [\bar\tau; \bar\zeta]$ such that $\bar\tau^2 = \bar\eta^T B\bar\eta \leq 1$, $\bar\zeta^T Q_j\bar\zeta = \bar\eta^T B_j\bar\eta \leq \vartheta^2\rho^2$, $1 \leq j \leq J$, and $\bar\eta^T A_{i,z}\bar\eta \geq \mathrm{SDP}_{i,z}$. Since $|\bar\tau| \leq 1$ and $\bar\zeta^T Q_j\bar\zeta \leq \vartheta^2\rho^2$, we have

$$\mathrm{Opt}_{i,z}(\vartheta\rho) \geq \bar\eta^T A_{i,z}\bar\eta \geq \mathrm{SDP}_{i,z} > -a_i[z],$$

as claimed.

We have arrived at the following result:

> *The AARC of an arbitrary uncertain LO problem, the uncertainty set being the intersection of J ellipsoids centered at the origin, is computationally tractable, provided $J = 1$, and admits safe tractable approximation, tight within the factor* $9.19\sqrt{\ln(J)}$ *when $J > 1$.*

In fact the above approach can be extended even slightly beyond just affine decision rules. Specifically, in the case of an uncertain LO we could allow for the adjustable "fixed recourse" variables x_j — those for which all the coefficients in the objective and the constraints of instances are certain — to be *quadratic* in $P_j\zeta$, and for the remaining "non-fixed recourse" adjustable variables to be affine in $P_j\zeta$. Indeed, this modification does not alter the structure of (5.3.22).

5.3.4 Illustration: the AARC of Multi-Period Inventory Affected by Uncertain Demand

We are about to illustrate the AARC methodology by its application to the simple multi-product multi-period inventory model presented in Example 5.1 (see also p. 211).

Building the AARC of (5.2.3). We first decide on the information base of the "actual decisions" — vectors w_t of replenishment orders of instants $t = 1, ..., N$. Assuming that the part of the uncertain data, (i.e., of the demand trajectory $\zeta = \zeta^N = [\zeta_1; ...; \zeta_N]$) that becomes known when the decision on w_t should be made is the vector $\zeta^{t-1} = [\zeta_1; ...; \zeta_{t-1}]$ of the demands in periods preceding time t, we introduce affine decision rules

$$w_t = \omega_t + \Omega_t\zeta^{t-1} \tag{5.3.25}$$

for the orders; here ω_t, Ω_t form the coefficients of the decision rules we are seeking.

The remaining variables in (5.2.3), with a single exception, are analysis variables, and we allow them to be arbitrary affine functions of the entire demand trajectory ζ^N:

$$\begin{array}{lll} x_t = \xi_t + \Xi_t\zeta^N,\ t = 2, ..., N+1 & \text{[states]} \\ y_t = \eta_t + H_t\zeta^N,\ t = 1, ..., N & \text{[upper bounds on } [x_t]_+] \\ z_t = \pi_t + \Pi_t\zeta^N,\ t = 1, ..., N & \text{[upper bounds on } [x_t]_-]. \end{array} \tag{5.3.26}$$

The only remaining variable C — the upper bound on the inventory management cost we intend to minimize — is considered as non-adjustable.

We now plug the affine decision rules in the objective and the constraints of (5.2.3), and require the resulting relations to be satisfied for all realizations of the uncertain data ζ^N from a given uncertainty set $\mathcal{Z}$, thus arriving at the AARC of our inventory model:

$$\begin{array}{l}
\text{minimize} \qquad C \\
\text{s.t. } \forall \zeta^N \in \mathcal{Z}: \\
C \geq \sum_{t=1}^N \left[c_{\text{h},t}^T[\eta_t + H_t\zeta^N] + c_{\text{b},t}^T[\pi_t + \Pi_t\zeta^N] + c_{\text{o},t}^T[\omega_t + \Omega_t\zeta^{t-1}] \right] \\
\xi_t + \Xi_t\zeta^N = \left\{ \begin{array}{l} \xi_{t-1} + \Xi_{t-1}\zeta^N + [\omega_t + \Omega_t\zeta^{t-1}] - \zeta_t,\ 2 \leq t \leq N \\ x_0 + \omega_1 - \zeta_1,\ t = 1 \end{array} \right. \\
\eta_t + H_t\zeta^N \geq 0,\ \eta_t + H_t\zeta^N \geq \xi_t + \Xi_t\zeta^N,\ 1 \leq t \leq N \\
\pi_t + \Pi_t\zeta^N \geq 0,\ \pi_t + \Pi_t\zeta^N \geq -\xi_t - \Xi_t\zeta^N,\ 1 \leq t \leq N \\
\underline{w}_t \leq \omega_t + \Omega_t\zeta^{t-1} \leq \overline{w}_t,\ 1 \leq t \leq N \\
q^T\left[\eta_t + H_t\zeta^N\right] \leq r
\end{array} \tag{5.3.27}$$

the variables being C and the coefficients $\omega_t, \Omega_t, ..., \pi_t, \Pi_t$ of the affine decision rules.

We see that the problem in question has fixed recourse (it always is so when the uncertainty affects just the constant terms in conic constraints) and is nothing but an explicit semi-infinite LO program. Assuming the uncertainty set $\mathcal{Z}$ to be computationally tractable, we can invoke Theorem 1.1 and reformulate this semi-infinite problem as a computationally tractable one. For example, with *box uncertainty*:

$$\mathcal{Z} = \{\zeta^N \in \mathbb{R}_+^{N\times d} : \underline{\zeta}_t \leq \zeta_t \leq \overline{\zeta}_t,\ 1 \leq t \leq N\},$$

the semi-infinite LO program (5.3.27) can be immediately rewritten as an explicit "certain" LO program. Indeed, after replacing the semi-infinite coordinate-wise vector inequalities/equations appearing in (5.3.27) by equivalent systems of scalar semi-infinite inequalities/equations and representing the semi-infinite linear equations by pairs of opposite semi-infinite linear inequalities, we end up with a semi-infinite optimization program with a certain linear objective and finitely many constraints of the form

$$\forall \left(\zeta_t^i \in [\underline{\zeta}_t^i, \overline{\zeta}_t^i], t \leq N, i \leq d\right) : p^\ell[y] + \sum_{i,t} \zeta_t^i p_{ti}^\ell[y] \leq 0$$

(ℓ is the serial number of the constraint, y is the vector comprised of the decision variables in (5.3.27), and $p^\ell[y]$, $p_{ti}^\ell[y]$ are given affine functions of y). The above semi-infinite constraint can be represented by a system of linear inequalities

$$\begin{array}{l}
\underline{\zeta}_t^i p_{ti}^\ell[y] \leq u_{ti}^\ell \\
\overline{\zeta}_t^i p_{ti}^\ell[y] \leq u_{ti}^\ell \\
p^\ell[y] + \sum_{t,i} u_{ti}^\ell \leq 0,
\end{array}$$

in variables y and additional variables u_{ti}^ℓ. Putting all these systems of inequalities together and augmenting the resulting system of linear constraints with our original objective to be minimized, we end up with an explicit LO program that is equivalent to (5.3.27).

Some remarks are in order:

1. We could act similarly when building the AARC of any uncertain LO problem with fixed recourse and "well-structured" uncertainty set, e.g., one given by an explicit polyhedral/conic quadratic/semidefinite representation. In the latter case, the resulting tractable reformulation of the AARC would be an explicit linear/conic quadratic/semidefinite program of sizes that are polynomial in the sizes of the instances and in the size of conic description of the uncertainty set. Moreover, the "tractable reformulation" of the AARC can be built automatically, by a kind of compilation.

2. Note how flexible the AARC approach is: we could easily incorporate additional constraints, (e.g., those forbidding backlogged demand, expressing lags in acquiring information on past demands and/or lags in executing the replenishment orders, etc.). Essentially, the only thing that matters is that we are dealing with an uncertain LO problem with fixed recourse. This is in sharp contrast with the ARC. As we have already mentioned, there is, essentially, only one optimization technique — Dynamic Programming — that with luck can be used to process the (general-type) ARC numerically. To do so, one needs indeed a lot of luck — to be "computationally tractable," Dynamic Programming imposes many highly "fragile" limitations on the structure and the sizes of instances. For example, the effort to solve the "true" ARC of our toy Inventory problem by Dynamic Programming blows up exponentially with the number of products d (we can say that $d = 4$ is already "too big"); in contrast to this, the AARC does not suffer of "curse of dimensionality" and scales reasonably well with problem's sizes.

3. Note that we have no difficulties processing uncertainty-affected *equality constraints* (such as state equations above) — this is something that we cannot afford with the usual — non-adjustable — RC (how could an equation remain valid when the variables are kept constant, and the coefficients are perturbed?).

4. Above, we "immunized" affine decision rules against uncertainty in the worst-case-oriented fashion — by requiring the constraints to be satisfied for *all* realizations of uncertain data from $\mathcal{Z}$. Assuming ζ to be random, we could replace the worst-case interpretation of the uncertain constraints with their chance constrained interpretation. To process the "chance constrained" AARC, we could use all the "chance constraint machinery" we have developed so far for the RC, exploiting the fact that for fixed recourse there is no essential difference between the structure of the RC and that of the AARC.

Of course, all the nice properties of the AARC we have just mentioned have their price — in general, as in our toy inventory example, we have no idea of how much we lose in terms of optimality when passing from general decision rules to affine rules. At present, we are not aware of any theoretical tools for evaluating such a loss. Moreover, it is easy to build examples showing that sticking to affine decision rules can indeed be costly; it even may happen that the AARC is infeasible, while the ARC is not. Much more surprising is the fact that there are meaningful situations where the AARC is unexpectedly good. Here we present a single simple example.

Consider our inventory problem in the single-product case with added constraints that no backlogged demand is allowed and that the amount of product in the inventory should remain between two given positive bounds. Assuming box uncertainty in the demand, the "true" ARC of the uncertain problem is well within the grasp of Dynamic Programming, and thus we can measure the "non-optimality" of affine decision rules experimentally — by comparing the optimal values of the true ARC with those of the AARC as well as of the non-adjustable RC. To this end, we generated at random several hundreds of data sets for the problem with time horizon

$N = 10$ and filtered out all data sets that led to infeasible ARC (it indeed can be infeasible due to the presence of upper and lower bounds on the inventory level and the fact that we forbid backlogged demand). We did our best to get as rich a family of examples as possible — those with time-independent and with time-dependent costs, various levels of demand uncertainty (from 10% to 50%), etc. We then solved ARCs, AARCs and RCs of the remaining "well-posed" problems — the ARCs by Dynamic Programming, the AARCs and RCs — by reduction to explicit LO programs. The number of "well-posed" problems we processed was 768, and the results were as follows:

1. To our great surprise, *in every one of the 768 cases we have analyzed, the computed optimal values of the "true" ARC and the AARC were identical.* Thus, there is an "experimental evidence" that in the case of our single-product inventory problem, the affine decision rules allow one to reach "true optimality."

   Quite recently, D. Bertsimas, D. Iancu and P. Parrilo have demonstrated [30] that the above "experimental evidence" has solid theoretical reasons, specifically, they have established the following remarkable and unexpected result:

   Consider the multi-stage uncertainty-affected decision making problem

$$\left\{\min_{C,x,w}\left\{C:\ \begin{array}{l} C \geq \sum_{t=1}^{N}[c_t w_t + h_t(x_t)] \\ x_t = \alpha_t x_{t-1} + \beta_t w_t + \gamma_t \zeta_t,\ 1 \leq t \leq N \\ \underline{w}_t \leq w_t \leq \overline{w}_t,\ 1 \leq t \leq N \end{array}\right\} : \zeta^N = [\zeta_1; ...; \zeta_N] \in \mathcal{Z} \subset \mathbb{R}^N\right\}$$

   with uncertain data $\zeta^N = [\zeta_1; ...; \zeta_N]$, the variables in the problem being C (non-adjustable), w_t (allowed to depend on the "past demands" ζ^{t-1}), $1 \leq t \leq N$, and x_t (fully adjustable – allowed to depend on ζ^N); here the functions $h_t(\cdot)$ are convex functions on the axis. Assume, further, that $\mathcal{Z}$ is a box, and consider the ARC and the AARC of the problem, that is, the infinite-dimensional problems

$$\min_{C,x(\cdot),w(\cdot),z(\cdot)}\left\{C:\ \begin{array}{l} C \geq \sum_{t=1}|^N[c_t w_t(\zeta^{t-1}) + z_t(\zeta^N)] \\ x_t(\zeta^N) = \alpha_t x_{t-1}(\zeta^N) + \beta_t w_t(\zeta^{t-1}) + \gamma_t \zeta_t,\ 1 \leq t \leq N \\ z_t(\zeta^N) \geq h_t(x_t(\zeta^N)), 1 \leq t \leq N \\ \underline{w}_t \leq w_t(\zeta^{t-1}) \leq \overline{w}_t, 1 \leq t \leq N \end{array}\right\}\ \forall \zeta^N \in \mathcal{Z}\right\}$$

   where the optimization is taken over arbitrary functions $x_t(\zeta^N)$, $w_t(\zeta^{t-1})$, $z_t(\zeta^N)$ (ARC) and over affine functions $x_t(\zeta^T$, $z_t(\zeta^t)$, $w_t(\zeta^{t-1})$ (AARC); here slacks $z_t(\cdot)$ are upper bounds on costs $h_t(x_t(\cdot))$. The the optimal solution to the AARC is an optimal solution to the ARC as well.

   Note that the single product version of our Inventory Management problem satisfies the premise of the latter result, provided that the uncertainty set is a box (as is the case in the experiments we have reported), and the corresponding functions $h_t(\cdot)$ are not only convex, but also piecewise linear, the domain of h_t being $x \leq r/q$; in this case what was called AARC of the problem, is nothing but our AARC (5.3.26) – (5.3.27) where we further restrict $z_t(\cdot)$ to be identical to $y_t(\cdot)$. Thus, in the single product case the AARC of the Inventory Management problem in question is *equivalent* to its ARC.

   Toe the best of our knowledge, the outlined result of Bertsimas, Iancu and Parrilo yields the only known for the time being generic example of a meaningful multi-stage uncertainty affected decision making problem where the affine decision rules are provably optimal. This remarkable result is very "fragile," e.g., it cannot be extended on multi-product inventory, or on the case when aside of bounds on replenishment orders in every period there are

Range of $\frac{\text{Opt(RC)}}{\text{Opt(AARC)}}$	1	$(1,2]$	$(2,10]$	$(10,1000]$	∞
Frequency in the sample	38%	23%	14%	11%	15%

Table 5.1: Experiments with ARCs, AARCs and RCs of randomly generated single-product inventory problems affected by uncertain demand.

bounds on cumulative replenishment orders, etc. It should be added that the phenomenon in question seems to be closely related to our intention to optimize the *guaranteed*, (i.e., the worst-case, w.r.t. demand trajectories from the uncertainty set), inventory management cost. When optimizing the "average" cost, the ARC frequently becomes significantly less expensive than the AARC.[2]

2. The (equal to each other) optimal values of the ARC and the AARC in many cases were much better than the optimal value of the RC, as it is seen from table 5.1. In particular, in 40% of the cases the RC was at least twice as bad in terms of the (worst-case) inventory management cost as the ARC/AARC, and in 15% of the cases the RC was in fact infeasible.

The bottom line is twofold. First, we see that in multi-stage decision making there exist meaningful situations where the AARC, while "not less computationally tractable" than the RC, is much more flexible and much less conservative. Second, the AARC is not necessarily "significantly inferior" as compared to the ARC.

5.4 Adjustable Robust Optimization and Synthesis of Linear Controllers

While the usefulness of affine decision rules seems to be heavily underestimated in the "OR-style multi-stage decision making," they play one of the central roles in Control. Our next goal is to demonstrate that the use of AARC can render important Control implications.

5.4.1 Robust Affine Control over Finite Time Horizon

Consider a discrete time linear dynamical system

$$\begin{array}{rcl} x_0 & = & z \\ x_{t+1} & = & A_t x_t + B_t u_t + R_t d_t \ , \ t = 0, 1, ... \\ y_t & = & C_t x_t + D_t d_t \end{array} \tag{5.4.1}$$

where $x_t \in \mathbb{R}^{n_x}$, $u_t \in \mathbb{R}^{n_u}$, $y_t \in \mathbb{R}^{n_y}$ and $d_t \in \mathbb{R}^{n_d}$ are the state, the control, the output and the exogenous input (disturbance) at time t, and A_t, B_t, C_t, D_t, R_t are known matrices of appropriate dimension.

[2]On this occasion, it is worthy of mention that affine decision rules were proposed many years ago, in the context of Multi-Stage Stochastic Programming, by A. Charnes. In Stochastic Programming, people are indeed interested in optimizing the expected value of the objective, and soon it became clear that in this respect, the affine decision rules can be pretty far from being optimal. As a result, the simple — and extremely useful from the computational perspective — concept of affine decision rules remained completely forgotten for many years.

Notational convention. Below, given a sequence of vectors $e_0, e_1, ...$ and an integer $t \geq 0$, we denote by e^t the initial fragment of the sequence: $e^t = [e_0; ...; e_t]$. When t is negative, e^t, by definition, is the zero vector.

Affine control laws. A typical problem of (finite-horizon) Linear Control associated with the "open loop" system (5.4.1) is to "close" the system by a non-anticipative affine output-based control law

$$u_t = g_t + \sum\nolimits_{\tau=0}^{t} G_{t\tau} y_\tau \tag{5.4.2}$$

(here the vectors g_t and matrices $G_{t\tau}$ are the parameters of the control law). The closed loop system (5.4.1), (5.4.2) is required to meet prescribed design specifications. We assume that these specifications are represented by a system of linear inequalities

$$Aw^N \leq b \tag{5.4.3}$$

on the *state-control trajectory* $w^N = [x_0; ...; x_{N+1}; u_0; ...; u_N]$ over a given finite time horizon $t = 0, 1, ..., N$.

An immediate observation is that for a given control law (5.4.2) the dynamics (5.4.1) specifies the trajectory as an affine function of the initial state z and the sequence of disturbances $d^N = (d_0, ..., d_N)$:

$$w^N = w_0^N[\gamma] + W^N[\gamma]\zeta, \ \zeta = (z, d^N),$$

where $\gamma = \{g_t, G_{t\tau}, 0 \leq \tau \leq t \leq N\}$, is the "parameter" of the underlying control law (5.4.2). Substituting this expression for w^N into (5.4.3), we get the following system of constraints on the decision vector γ:

$$A\left[w_0^N[\gamma] + W^N[\gamma]\zeta\right] \leq b. \tag{5.4.4}$$

If the disturbances d^N and the initial state z are certain, (5.4.4) is "easy" — it is a system of constraints on γ with certain data. Moreover, in the case in question we lose nothing by restricting ourselves with "off-line" control laws (5.4.2) — those with $G_{t\tau} \equiv 0$; when restricted onto this subspace, let it be called Γ, in the γ space, the function $w_0^N[\gamma] + W^N[\gamma]\zeta$ turns out to be bi-affine in γ and in ζ, so that (5.4.4) reduces to a system of explicit linear inequalities on $\gamma \in \Gamma$. Now, when the disturbances and/or the initial state are *not* known in advance, (which is the only case of interest in Robust Control), (5.4.4) becomes an uncertainty-affected system of constraints, and we could try to solve the system in a robust fashion, e.g., to seek a solution γ that makes the constraints feasible for all realizations of $\zeta = (z, d^N)$ from a given uncertainty set $\mathcal{ZD}^N$, thus arriving at the system of semi-infinite scalar constraints

$$A\left[w_0^N[\gamma] + W^N[\gamma]\zeta\right] \leq b \ \forall \zeta \in \mathcal{ZD}^N. \tag{5.4.5}$$

Unfortunately, the semi-infinite constraints in this system are *not* bi-affine, since the dependence of w_0^N, W^N on γ is highly nonlinear, unless γ is restricted to vary in Γ. Thus, when seeking "on-line" control laws (those where $G_{t\tau}$ can be nonzero), (5.4.5) becomes a system of highly nonlinear semi-infinite constraints and as such seems to be severely computationally intractable (the feasible set corresponding to (5.4.4) can be in fact nonconvex). One possibility to circumvent this difficulty would be to switch from control laws that are affine in the outputs y_t to those affine in disturbances and the initial state (cf. approach of [51]). This, however, could be problematic in the situations when we do not observe z and d_t directly. The good news is that we can overcome this difficulty without requiring d_t and z to be observable, the remedy being a suitable re-parameterization of affine control laws.

5.4.2 Purified-Output-Based Representation of Affine Control Laws and Efficient Design of Finite-Horizon Linear Controllers

Imagine that in parallel with controlling (5.4.1) with the aid of a non-anticipating output-based control law $u_t = U_t(y_0, ..., y_t)$, we run the *model* of (5.4.1) as follows:

$$\begin{array}{rcl} \widehat{x}_0 & = & 0 \\ \widehat{x}_{t+1} & = & A_t\widehat{x}_t + B_t u_t \\ \widehat{y}_t & = & C_t\widehat{x}_t \\ v_t & = & y_t - \widehat{y}_t. \end{array} \tag{5.4.6}$$

Since we know past controls, we can run this system in an "on-line" fashion, so that the *purified output* v_t becomes known when the decision on u_t should be made. An immediate observation is that *the purified outputs are completely independent of the control law in question — they are affine functions of the initial state and the disturbances $d_0, ..., d_t$, and these functions are readily given by the dynamics of* (5.4.1).

Indeed, from the descriptions of the open-loop system and the model, it follows that the differences $\delta_t = x_t - \widehat{x}_t$ evolve with time according to the equations

$$\begin{array}{rcl} \delta_0 & = & z \\ \delta_{t+1} & = & A_t + R_t d_t,\ t = 0, 1, ... \end{array}$$

while

$$v_t = C_t\delta_t + D_t d_t.$$

From these relations it follows that

$$v_t = \mathcal{V}_t^d d^t + \mathcal{V}_t^z z \tag{5.4.7}$$

with matrices $\mathcal{V}_t^d$, $\mathcal{V}_t^z$ depending solely on the matrices $A_\tau, B_\tau, ...$, $0 \le \tau \le t$, and readily given by these matrices.

Now, it was mentioned that $v_0, ..., v_t$ are known when the decision on u_t should be made, so that we can consider *purified-output-based* (POB) affine control laws

$$u_t = h_t + \sum\nolimits_{\tau=0}^{t} H_{t\tau} v_\tau.$$

The complete description of the dynamical system "closed" by this control is

$$\begin{array}{|l|} \hline \underline{\text{plant:}} \\ (a): \left\{ \begin{array}{rcl} x_0 & = & z \\ x_{t+1} & = & A_t x_t + B_t u_t + R_t d_t \\ y_t & = & C_t x_t + D_t d_t \end{array} \right. \\ \hline \underline{\text{model:}} \\ (b): \left\{ \begin{array}{rcl} \widehat{x}_0 & = & 0 \\ \widehat{x}_{t+1} & = & A_t\widehat{x}_t + B_t u_t \\ \widehat{y}_t & = & C_t\widehat{x}_t \end{array} \right. \\ \hline \underline{\text{purified outputs:}} \\ \underline{(c): \quad v_t \ = \ y_t - \widehat{y}_t} \\ \hline \underline{\text{control law:}} \\ (d): \quad u_t \ = \ h_t + \sum\limits_{\tau=0}^{t} H_{t\tau} v_\tau \\ \hline \end{array} \tag{5.4.8}$$

The main result. We are about to prove the following simple and fundamental fact:

Theorem 5.3

(i) *For every affine control law in the form of* (5.4.2)*, there exists a control law in the form of* (5.4.8.d) *that, whatever be the initial state and a sequence of inputs, results in exactly the same state-control trajectories of the closed loop system;*

(ii) *Vice versa, for every affine control law in the form of* (5.4.8.d)*, there exists a control law in the form of* (5.4.2) *that, whatever be the initial state and a sequence of inputs, results in exactly the same state-control trajectories of the closed loop system;*

(iii) [bi-affinity] *The state-control trajectory* w^N *of closed loop system* (5.4.8) *is affine in* z, d^N *when the parameters* $\eta = \{h_t, H_{t\tau}\}_{0\leq\tau\leq t\leq N}$ *of the underlying control law are fixed, and is affine in* η *when* z, d^N *are fixed:*

$$w^N = \omega[\eta] + \Omega_z[\eta]z + \Omega_d[\eta]d^N \tag{5.4.9}$$

for some vectors $\omega[\eta]$ *and matrices* $\Omega_z[\eta]$, $\Omega_d[\eta]$ *depending affinely on* η.

Proof. (i): Let us fix an affine control law in the form of (5.4.2), and let $x_t = X_t(z, d^{t-1})$, $u_t = U_t(z, d^t)$, $y_t = Y_t(z, d^t)$, $v_t = V_t(z, d^t)$ be the corresponding states, controls, outputs, and purified outputs. To prove (i) it suffices to show that for every $t \geq 0$ with properly chosen vectors q_t and matrices $Q_{t\tau}$ one has

$$\forall (z, d^t) : Y_t(z, d^t) = q_t + \sum_{\tau=0}^{t} Q_{t\tau} V_\tau(z, d^\tau). \tag{I$_t$}$$

Indeed, given the validity of these relations and taking into account (5.4.2), we would have

$$U_t(z, d^t) \equiv g_t + \sum_{\tau=0}^{t} G_{t\tau} Y_\tau(z, d^\tau) \equiv h_t + \sum_{\tau=0}^{t} H_{t\tau} V(z, d^\tau) \tag{II$_t$}$$

with properly chosen h_t, $H_{t\tau}$, so that the control law in question can indeed be represented as a linear control law via purified outputs.

We shall prove (I$_t$) by induction in t. The base $t = 0$ is evident, since by (5.4.8.a–c) we merely have $Y_0(z, d^0) \equiv V_0(z, d^0)$. Now let $s \geq 1$ and assume that relations (I$_t$) are valid for $0 \leq t < s$. Let us prove the validity of (I$_s$). From the validity of (I$_t$), $t < s$, it follows that the relations (II$_t$), $t < s$, take place, whence, by the description of the model system, $\widehat{x}_s = \widehat{X}_s(z, d^{s-1})$ is affine in the purified outputs, and consequently the same is true for the model outputs $\widehat{y}_s = \widehat{Y}_s(z, d^{s-1})$:

$$\widehat{Y}_s(z, d^{s-1}) = p_s + \sum_{\tau=0}^{s-1} P_{s\tau} V_\tau(z, d^\tau).$$

We conclude that with properly chosen p_s, $P_{s\tau}$ we have

$$Y_s(z, d^s) \equiv \widehat{Y}_s(z, d^{s-1}) + V_s(z, d^s) = p_s + \sum_{\tau=0}^{s-1} P_{s\tau} V_\tau(z, d^\tau) + V_s(z, d^s),$$

as required in (I$_s$). Induction is completed, and (i) is proved.

(ii): Let us fix a linear control law in the form of (5.4.8.d), and let $x_t = X_t(z, d^{t-1})$, $\widehat{x}_t = \widehat{X}_t(z, d^{t-1})$, $u_t = U_t(z, d^t)$, $y_t = Y_t(z, d^t)$, $v_t = V_t(z, d^t)$ be the corresponding actual and model states, controls, and actual and purified outputs. We should verify that the state-control dynamics in question can be obtained from an appropriate control law in the form of (5.4.2). To this end, similarly to the proof of (i), it suffices to show that for every $t \geq 0$ one has

$$V_t(z, d^t) \equiv q_t + \sum_{\tau=0}^{t} Q_{t\tau} Y_\tau(z, d^\tau) \tag{III$_t$}$$

with properly chosen q_t, $Q_{t\tau}$. We again apply induction in t. The base $t = 0$ is again trivially true due to $V_0(z, d^0) \equiv Y_0(z, d^0)$. Now let $s \geq 1$, and assume that relations (III$_t$) are valid for $0 \leq t < s$, and let us prove that (III$_s$) is valid as well. From the validity of (III$_t$), $t < s$, and from (5.4.8.d) it follows that

$$t < s \Rightarrow U_t(z, d^t) = c_t + \sum_{\tau=0}^{t} C_{t\tau} Y_\tau(z, d^\tau)$$

with properly chosen c_t and $C_{t\tau}$. From these relations and the description of the model system it follows that its state $\widehat{X}_s(z, d^{s-1})$ at time s, and therefore the model output $\widehat{Y}_s(z, d^{s-1})$, are affine functions of $Y_0(z, d^0)$,..., $Y_{s-1}(z, d^{s-1})$:

$$\widehat{Y}_s(z, d^{s-1}) = p_s + \sum_{\tau=0}^{s-1} P_{s\tau} Y_\tau(z, d^\tau)$$

with properly chosen p_s, $P_{s\tau}$. It follows that

$$V_s(z, d^s) \equiv Y_s(z, d^s) - \widehat{Y}_s(z, d^{s-1}) = Y_s(z, d^s) - p_s - \sum_{\tau=0}^{s-1} P_{s\tau} Y_\tau(z, d^\tau),$$

as required in (III$_s$). Induction is completed, and (ii) is proved.

(iii): For $0 \leq s \leq t$ let

$$A_s^t = \begin{cases} \prod_{r=s}^{t-1} A_r, & s < t \\ I, & s = t \end{cases}$$

Setting $\delta_t = x_t - \widehat{x}_t$, we have by (5.4.8.$a$–$b$)

$$\delta_{t+1} = A_t \delta_t + R_t d_t, \; \delta_0 = z \Rightarrow \delta_t = A_0^t z + \sum_{s=0}^{t-1} A_{s+1}^t R_s d_s$$

(from now on, sums over empty index sets are zero), whence

$$v_\tau = C_\tau \delta_\tau + D_\tau d_\tau = C_\tau A_0^\tau z + \sum_{s=0}^{\tau-1} C_\tau A_{s+1}^\tau R_s d_s + D_\tau d_\tau. \tag{5.4.10}$$

Therefore control law (5.4.8.d) implies that

$$\begin{array}{rcl} u_t & = & h_t + \sum\limits_{\tau=0}^{t} H_{t\tau} v_\tau = \underbrace{h_t}_{\nu_t[\eta]} + \underbrace{\left[\sum\limits_{\tau=0}^{t} H_{t\tau} C_\tau A_0^\tau\right]}_{N_t[\eta]} z \\ & & + \sum\limits_{s=0}^{t-1} \underbrace{\left[H_{ts} D_s + \sum\limits_{\tau=s+1}^{t} H_{t\tau} C_\tau A_{s+1}^\tau R_s\right]}_{N_{ts}[\eta]} d_s + \underbrace{H_{tt} D_t}_{N_{tt}[\eta]} d_t \\ & = & \nu_t[\eta] + N_t[\eta] z + \sum\limits_{s=0}^{t} N_{ts}[\eta] d_s, \end{array} \tag{5.4.11}$$

whence, invoking (5.4.8.a),

$$\begin{array}{rcl} x_t & = & A_0^t z + \sum\limits_{\tau=0}^{t-1} A_{\tau+1}^t [B_\tau u_\tau + R_\tau d_\tau] = \underbrace{\left[\sum\limits_{\tau=0}^{t-1} A_{\tau+1}^t B_\tau h_t\right]}_{\mu_t[\eta]} \\ & & + \underbrace{\left[A_0^t + \sum\limits_{\tau=0}^{t-1} A_{\tau+1}^t B_\tau N_\tau[\eta]\right]}_{M_t[\eta]} z \\ & & + \sum\limits_{s=0}^{t-1} \underbrace{\left[\sum\limits_{\tau=s}^{t-1} A_{\tau+1}^t B_\tau N_{\tau s}[\eta] + A_{s+1}^t B_s R_s\right]}_{M_{ts}[\eta]} d_s \\ & = & \mu_t[\eta] + M_t[\eta] z + \sum\limits_{s=0}^{t-1} M_{ts}[\eta] d_s. \end{array} \tag{5.4.12}$$

We see that the states x_t, $0 \le t \le N+1$, and the controls u_t, $0 \le t \le N$, of the closed loop system (5.4.8) are affine functions of z, d^N, and the corresponding "coefficients" $\mu_t[\eta]$,...,$N_{ts}[\eta]$ are affine vector- and matrix-valued functions of the parameters $\eta = \{h_t, H_{t\tau}\}_{0 \le \tau \le t \le N}$ of the underlying control law (5.4.8.d). □

The consequences. The representation (5.4.8.d) of affine control laws is incomparably better suited for design purposes than the representation (5.4.2), since, as we know from Theorem 5.3.(iii), *with controller* (5.4.8.d), *the state-control trajectory* w^N *becomes bi-affine in* $\zeta = (z, d^N)$ *and in the parameters* $\eta = \{h_t, H_{t\tau}, 0 \le \tau \le t \le N\}$ *of the controller:*

$$w^N = \omega^N[\eta] + \Omega^N[\eta]\zeta \tag{5.4.13}$$

with vector- and matrix-valued functions $\omega^N[\eta]$, $\Omega^N[\eta]$ *affinely depending on* η *and readily given by the dynamics* (5.4.1). Substituting (5.4.13) into (5.4.3), we arrive at the system of semi-infinite *bi-affine* scalar inequalities

$$A\left[\omega^N[\eta] + \Omega^N[\eta]\zeta\right] \le b \tag{5.4.14}$$

in variables η, and can use the tractability results from lectures 1, 4 in order to solve efficiently the RC/GRC of this uncertain system of scalar linear constraints. For example, we can process efficiently the GRC setting of the semi-infinite constraints (5.4.13)

$$\begin{array}{r} a_i^T\left[\omega^N[\eta] + \Omega^N[\eta][z; d^N]\right] - b_i \le \alpha_i^z \mathrm{dist}(z, \mathcal{Z}) + \alpha_d^i \mathrm{dist}(d^N, \mathcal{D}^N) \\ \forall [z; d^N]\ \forall i = 1, ..., I \end{array} \tag{5.4.15}$$

where $\mathcal{Z}$, $\mathcal{D}^N$ are "good," (e.g., given by strictly feasible semidefinite representations), closed convex normal ranges of z, d^N, respectively, and the distances are defined via the $\|\cdot\|_\infty$ norms (this setting corresponds to the "structured" GRC, see Definition 4.3). By the results of section 4.3, system (5.4.15) is equivalent to the system of constraints

$$\begin{array}{l} \forall(i, 1 \leq i \leq I): \\ (a) \quad a_i^T \left[\omega^N[\eta] + \Omega^N[\eta][z; d^N]\right] - b_i \leq 0 \ \forall [z; d^N] \in \mathcal{Z} \times \mathcal{D}^N \\ (b) \quad \|a_i^T \Omega_z^N[\eta]\|_1 \leq \alpha_z^i \qquad (c) \quad \|a_i^T \Omega_d^N[\eta]\|_1 \leq \alpha_d^i, \end{array} \tag{5.4.16}$$

where $\Omega^N[\eta] = \left[\Omega_z^N[\eta], \Omega_d^N[\eta]\right]$ is the partition of the matrix $\Omega^N[\eta]$ corresponding to the partition $\zeta = [z; d^N]$. Note that in (5.4.16), the semi-infinite constraints (a) admit explicit semidefinite representations (Theorem 1.1), while constraints $(b-c)$ are, essentially, just linear constraints on η and on α_z^i, α_d^i. As a result, (5.4.16) can be thought of as a computationally tractable system of convex constraints on η *and on the sensitivities* α_z^i, α_d^i, and we can minimize under these constraints a "nice," (e.g., convex), function of η and the sensitivities. Thus, after passing to the POB representation of affine control laws, we can process efficiently specifications expressed by systems of linear inequalities, to be satisfied in a robust fashion, on the (finite-horizon) state-control trajectory.

The just summarized nice consequences of passing to the POB control laws are closely related to the tractability of AARCs of uncertain LO problems with fixed recourse, specifically, as follows. Let us treat the state equations (5.4.1) coupled with the design specifications (5.4.3) as a system of uncertainty-affected linear constraints on the state-control trajectory w, the uncertain data being $\zeta = [z; d^N]$. Relations (5.4.10) say that the purified outputs v_t are known in advance, completely independent of what the control law in use is, *linear* functions of ζ. With this interpretation, a POB control law becomes a collection of affine decision rules that specify the decision variables u_t as affine functions of $P_t\zeta \equiv [v_0; v_1; ...; v_t]$ and simultaneously, via the state equations, specify the states x_t as affine functions of $P_{t-1}\zeta$. Thus, when looking for a POB control law that meets our design specifications in a robust fashion, we are doing nothing but solving the RC (or the GRC) of an uncertain LO problem in affine decision rules possessing a prescribed "information base." On closest inspection, this uncertain LO problem is with fixed recourse, and therefore its robust counterparts are computationally tractable.

Remark 5.2 *It should be stressed that the re-parameterization of affine control laws underlying Theorem 5.3 (and via this Theorem — the nice tractability results we have just mentioned) is nonlinear. As a result, it can be of not much use when we are optimizing over affine control laws satisfying additional restrictions rather than over all affine control laws.*

Assume, e.g., that we are seeking control in the form of a simple output-based linear feedback:

$$u_t = G_t y_t.$$

This requirement is just a system of simple linear constraints on the parameters of the control law in the form of (5.4.2), which, however, does not help much, since, as we have already explained, optimization over control laws in this form is by itself difficult. And when passing to affine control laws in the form of (5.4.8.d), the requirement that our would-be control should be a linear output-based feedback becomes a system of highly nonlinear constraints on our new design parameters η, and the synthesis again turns out to be difficult.

Example: Controlling finite-horizon gains. Natural design specification pertaining to finite-horizon Robust Linear Control are in the form of bounds on finite-horizon *gains* $\mathrm{z2x}^N$, $\mathrm{z2u}^N$, $\mathrm{d2x}^N$, $\mathrm{d2u}^N$ defined as follows: with a linear, (i.e., with $h_t \equiv 0$) control law (5.4.8.d), the states x_t and the controls u_t are linear functions of z and d^N:

$$x_t = X_t^z[\eta]z + X_t^d[\eta]d^N,\ u_t = U_t^z[\eta]z + U_t^d[\eta]d^N$$

with matrices $X_t^z[\eta]$,...,$U_t^d[\eta]$ affinely depending on the parameters η of the control law. Given t, we can define the z to x_t *gains* and the *finite-horizon* z to x gain as $\mathrm{z2x}_t(\eta) = \max\limits_z\{\|X_t^z[\eta]z\|_\infty : \|z\|_\infty \leq 1\}$ and $\mathrm{z2x}^N(\eta) = \max\limits_{0\leq t\leq N} \mathrm{z2x}_t(\eta)$. The definitions of the z to u gains $\mathrm{z2u}_t(\eta)$, $\mathrm{z2u}^N(\eta)$ and the "disturbance to x/u" gains $\mathrm{d2x}_t(\eta)$, $\mathrm{d2x}^N(\eta)$, $\mathrm{d2u}_t(\eta)$, $\mathrm{d2u}^N(\eta)$ are completely similar, e.g., $\mathrm{d2u}_t(\eta) = \max\limits_{d^N}\{\|U_t^d[\eta]d^N\|_\infty : \|d^N\|_\infty \leq 1\}$ and $\mathrm{d2u}^N(\eta) = \max\limits_{0\leq t\leq N} \mathrm{d2u}_t(\eta)$. The finite-horizon gains clearly are nonincreasing functions of the time horizon N and have a transparent Control interpretation; e.g., $\mathrm{d2x}^N(\eta)$ ("peak to peak d to x gain") is the largest possible perturbation in the states x_t, $t = 0, 1, ..., N$ caused by a unit perturbation of the sequence of disturbances d^N, both perturbations being measured in the $\|\cdot\|_\infty$ norms on the respective spaces. Upper bounds on ^{N}-gains (and on *global* gains like $\mathrm{d2x}^\infty(\eta) = \sup_{N\geq 0} \mathrm{d2x}^N(\eta)$) are natural Control specifications. With our purified-output-based representation of linear control laws, the finite-horizon specifications of this type result in explicit systems of linear constraints on η and thus can be processed routinely via LO. For example, an upper bound $\mathrm{d2x}^N(\eta) \leq \lambda$ on $\mathrm{d2x}^N$ gain is equivalent to the requirement $\sum_j |(X_t^d[\eta])_{ij}| \leq \lambda$ for all i and all $t \leq N$; since X_t^d is affine in η, this is just a system of linear constraints on η and on appropriate slack variables. Note that imposing bounds on the gains can be interpreted as passing to the GRC (5.4.15) in the case where the "desired behavior" merely requires $w^N = 0$, and the normal ranges of the initial state and the disturbances are the origins in the corresponding spaces: $\mathcal{Z} = \{0\}$, $\mathcal{D}^N = \{0\}$.

Non-affine control laws

So far, we focused on synthesis of finite-horizon *affine* POB controllers. Acting in the spirit of section 5.3.2, we can handle also synthesis of *quadratic* POB control laws — those where every entry of u_t, instead of being affine in the purified outputs $v^t = [v_0; ...; v_t]$, is allowed to be a quadratic function of v^t. Specifically, assume that we want to "close" the open loop system (5.4.1) by a non-anticipating control law in order to ensure that the state-control trajectory w^N of the closed loop system satisfies a given system S of linear constraints in a robust fashion, that is, for all realizations of the "uncertain data" $\zeta = [z; d^N]$ from a given uncertainty set $\mathcal{Z}_\rho^N = \rho\mathcal{Z}^N$ ($\rho > 0$ is, as always, the uncertainty level, and $\mathcal{Z} \ni 0$ is a closed convex set of "uncertain data of magnitude ≤ 1"). Let us use a *quadratic* POB control law in the form of

$$u_t^i = h_{it}^0 + h_{i,t}^T v^t + \frac{1}{\rho}[v^t]^T H_{i,t} v^t, \tag{5.4.17}$$

where u_t^i is i-th coordinate of the vector of controls at instant t, and h_{it}^0, h_{it} and H_{it} are, respectively, real, vector, and matrix parameters of the control law.[3] On a finite time horizon $0 \leq t \leq N$, such a quadratic control law is specified by ρ and the finite-dimensional vector $\eta = \{h_{it}^0, h_{it}, H_{it}\}_{\substack{1\leq i\leq \dim u\\ 0\leq t\leq N}}$. Now note that the purified outputs are well defined for any non-anticipating control law, not necessary affine, and they are *independent of the control law linear*

[3]The specific way in which the uncertainty level ρ affects the controls is convenient technically and is of no practical importance, since "in reality" the uncertainty level is a known constant.

functions of $\zeta^t \equiv [z; d^t]$. *The coefficients of these linear functions are readily given by the data* $A_\tau, ..., D_\tau$, $0 \le \tau \le t$ (see (5.4.7)). With this in mind, we see that *the controls, as given by* (5.4.17), *are quadratic functions of the initial state and the disturbances, the coefficients of these quadratic functions being affine in the vector* η *of parameters of our quadratic control law:*

$$u_t^i = \mathcal{U}_{it}^{(0)}[\eta] + [z; d^t]^T \mathcal{U}_{it}^{(1)}[\eta] + \frac{1}{\rho}[z; d^t]^T \mathcal{U}_{it}^{(2)}[\eta][z; d^t] \tag{5.4.18}$$

with *affine in* η reals/vectors/matrices $\mathcal{U}_{it}^{(\kappa)}[\eta]$, $\kappa = 0, 1, 2$. Plugging these representations of the controls into the state equations of the open loop system (5.4.1), we conclude that the states x_t^j of the closed loop system obtained by "closing" (5.4.1) by the quadratic control law (5.4.17), have the same "affine in η, quadratic in $[z; d^t]$" structure as the controls:

$$x_t^i = \mathcal{X}_{jt}^{(0)}[\eta] + [z; d^{t-1}]^T \mathcal{X}_{jt}^{(1)}[\eta] + \frac{1}{\rho}[z; d^{t-1}]^T \mathcal{X}_{jt}^{(2)}[\eta][z; d^{t-1}] \tag{5.4.19}$$

with affine in η reals/vectors/matrices $\mathcal{X}_{jt}^{(\kappa)}$, $\kappa = 0, 1, 2$.

Plugging representations (5.4.18), (5.4.19) into the system S of our target constraints, we end up with a system of semi-infinite constraints on the parameters η of the control law, specifically, the system

$$a_k[\eta] + 2\zeta^T p_k[\eta] + \frac{1}{\rho}\zeta^T R_k[\eta]\zeta \le 0 \,\forall \zeta = [z; d^N] \in \mathcal{Z}_\rho^N = \rho\mathcal{Z}^N, \; k = 1, ..., K, \tag{5.4.20}$$

where $a_k[\eta]$, $p_k[\eta]$ and $R_k[\zeta]$ are affine in η. Setting $P_k[\eta] = \left[\begin{array}{c|c} & p_k^T[\eta] \\ \hline p_k[\eta] & R_k[\eta] \end{array}\right]$, $\widehat{\zeta}_\rho[\zeta] = \left[\begin{array}{c|c} & \zeta^T \\ \hline \zeta & \zeta\zeta^T \end{array}\right]$ and denoting by $\widehat{\mathcal{Z}}_\rho^N$ the convex hull of the image of the set $\mathcal{Z}_\rho^N$ under the mapping $\zeta \mapsto \widehat{\zeta}_\rho[\zeta]$, system (5.4.20) can be rewritten equivalently as

$$a_k[\eta] + \mathrm{Tr}(P_k[\eta]\widehat{\zeta}) \le 0 \;\forall (\widehat{\zeta} \in \widehat{\mathcal{Z}}_\rho^N \equiv \rho\widehat{\mathcal{Z}}_1^N, k = 1, ..., K) \tag{5.4.21}$$

and we end up with a system of semi-infinite bi-affine scalar inequalities. From the results of section 5.3.2 it follows that this semi-infinite system:

- is computationally tractable, provided that $\mathcal{Z}^N$ is an ellipsoid $\{\zeta : \zeta^T Q\zeta \le 1\}$, $Q \succ 0$. Indeed, here $\widehat{\mathcal{Z}}_1^N$ is the semidefinite representable set

$$\{\left[\begin{array}{c|c} & \omega^T \\ \hline \omega & \Omega \end{array}\right] : \left[\begin{array}{c|c} 1 & \omega^T \\ \hline \omega & \Omega \end{array}\right] \succeq 0, \mathrm{Tr}(\Omega Q) \le 1\};$$

- admits a safe tractable approximation tight within the factor $\vartheta = O(1)\ln(J+1)$, provided that $\mathcal{Z}^N$ is the $\cap$-ellipsoidal uncertainty set $\{\zeta : \zeta^T Q_j \zeta \le 1, \, 1 \le j \le J\}$, where $Q_j \succeq 0$ and $\sum_j Q_j \succ 0$. This approximation is obtained when replacing the "true" uncertainty set $\widehat{\mathcal{Z}}_\rho^N$ with the semidefinite representable set

$$\mathcal{W}_\rho = \rho\{\left[\begin{array}{c|c} & \omega^T \\ \hline \omega & \Omega \end{array}\right] : \left[\begin{array}{c|c} 1 & \omega^T \\ \hline \omega & \Omega \end{array}\right] \succeq 0, \mathrm{Tr}(\Omega Q_j) \le 1, 1 \le j \le J\}$$

(recall that $\widehat{\mathcal{Z}}_\rho^N \subset \mathcal{W}_\rho \subset \widehat{\mathcal{Z}}_{\vartheta\rho}^N$).

5.4.3 Handling Infinite-Horizon Design Specifications

One might think that the outlined reduction of (discrete time) Robust Linear Control problems to Convex Programming, based on passing to the POB representation of affine control laws and deriving tractable reformulations of the resulting semi-infinite bi-affine scalar inequalities is intrinsically restricted to the case of finite-horizon control specifications. In fact our approach is well suited for handling infinite-horizon specifications — those imposing restrictions on the asymptotic behavior of the closed loop system. Specifications of the latter type usually have to do with the *time-invariant* open loop system (5.4.1):

$$\begin{array}{rcl} x_0 &=& z \\ x_{t+1} &=& Ax_t + Bu_t + Rd_t \ , \ t = 0, 1, ... \\ y_t &=& Cx_t + Dd_t \end{array} \tag{5.4.22}$$

From now on we assume that *the open loop system* (5.4.22) *is stable*, that is, the spectral radius of A is < 1 (in fact this restriction can be somehow circumvented, see below). Imagine that we "close" (5.4.22) by a *nearly time-invariant* POB control law *of order* k, that is, a law of the form

$$u_t = h_t + \sum_{s=0}^{k-1} H_s^t v_{t-s}, \tag{5.4.23}$$

where $h_t = 0$ for $t \geq N_*$ and $H_\tau^t = H_\tau$ for $t \geq N_*$ for a certain *stabilization time* N_*. From now on, all entities with negative indices are set to 0. While the "time-varying" part $\{h_t, H_\tau^t, 0 \leq t < N_*\}$ of the control law can be used to adjust the finite-horizon behavior of the closed loop system, its asymptotic behavior is as if the law were time-invariant: $h_t \equiv 0$ and $H_\tau^t \equiv H_\tau$ for all $t \geq 0$. Setting $\delta_t = x_t - \widehat{x}_t$, $H^t = [H_0^t, ..., H_{k-1}^t]$, $H = [H_0, ..., H_{k-1}]$, the dynamics (5.4.22), (5.4.6), (5.4.23) is given by

$$\begin{array}{rcl}
\overbrace{\left[\begin{array}{c} x_{t+1} \\ \delta_{t+1} \\ \delta_t \\ \vdots \\ \delta_{t-k+2} \end{array}\right]}^{\omega_{t+1}} &=& \overbrace{\left[\begin{array}{c|cccc} A & BH_0^tC & BH_1^tC & \dots & BH_{k-1}^tC \\ \hline & A & & & \\ & & A & & \\ & & & \ddots & \\ & & & & A \end{array}\right]}^{A_+[H^t]} \omega_t \\
&& + \overbrace{\left[\begin{array}{c|cccc} R & BH_0^tD & BH_1^tD & \dots & BH_{k-1}^tD \\ \hline & R & & & \\ & & R & & \\ & & & \ddots & \\ & & & & R \end{array}\right]}^{R_+[H^t]} \left[\begin{array}{c} d_t \\ d_t \\ d_{t-1} \\ \vdots \\ d_{t-k+1} \end{array}\right] \\
&& + \left[\begin{array}{c} Bh_t \\ 0 \\ \vdots \\ 0 \end{array}\right], \ t = 0, 1, 2, ..., \\
u_t &=& h_t + \sum_{\nu=0}^{k-1} H_\nu^t [C\delta_{t-\nu} + Dd_{t-\nu}].
\end{array} \tag{5.4.24}$$

We see that starting with time N_*, dynamics (5.4.24) is exactly as if the underlying control law were the time invariant POB law with the parameters $h_t \equiv 0$, $H^t \equiv H$. Moreover, since A is stable, we see that *system* (5.4.24) *is stable independently of the parameter* H *of the control* *law, and the resolvent* $\mathcal{R}_H(s) := (sI - A_+[H])^{-1}$ *of* $A_+[H]$ *is the affine in* H *matrix*

$$\left[\begin{array}{c|c|c|c|c} \mathcal{R}_A(s) & \mathcal{R}_A(s)BH_0C\mathcal{R}_A(s) & \mathcal{R}_A(s)BH_1C\mathcal{R}_A(s) & \ldots & \mathcal{R}_A(s)BH_{k-1}C\mathcal{R}_A(s) \\ \hline & \mathcal{R}_A(s) & & & \\ \hline & & \mathcal{R}_A(s) & & \\ \hline & & & \ddots & \\ \hline & & & & \mathcal{R}_A(s) \end{array}\right], \tag{5.4.25}$$

where $\mathcal{R}_A(s) = (sI - A)^{-1}$ *is the resolvent of* A.

Now imagine that the sequence of disturbances d_t is of the form $d_t = s^t d$, where $s \in \mathbb{C}$ differs from 0 and from the eigenvalues of A. From the stability of (5.4.24) it follows that as $t \to \infty$, the solution ω_t of the system, independently of the initial state, approaches the "steady-state" solution $\widehat{\omega}_t = s^t\mathcal{H}(s)d$, where $\mathcal{H}(s)$ is certain matrix. In particular, the state-control vector $w_t = \begin{bmatrix} x_t \\ u_t \end{bmatrix}$ approaches, as $t \to \infty$, the trajectory $\widehat{w}_t = s^t\mathcal{H}_{xu}(s)d$. The associated *disturbance-to-state/control transfer matrix* $\mathcal{H}_{xu}(s)$ is easily computable:

$$\mathcal{H}_{xu}(s) = \left[\begin{array}{c} \overbrace{\mathcal{R}_A(s)\left[R + \sum_{\nu=0}^{k-1} s^{-\nu}BH_\nu\left[D + C\mathcal{R}_A(s)R\right]\right]}^{\mathcal{H}_x(s)} \\ \hline \underbrace{\left[\sum_{\nu=0}^{k-1} s^{-\nu}H_\nu\right]\left[D + C\mathcal{R}_A(s)R\right]}_{\mathcal{H}_u(s)} \end{array}\right]. \tag{5.4.26}$$

The crucial fact is *that the transfer matrix* $\mathcal{H}_{xu}(s)$ *is affine in the parameters* $H = [H_0, ..., H_{k-1}]$ *of the nearly time invariant control law* (5.4.23). As a result, *design specifications representable as explicit convex constraints on the transfer matrix* $\mathcal{H}_{xu}(s)$ (these are typical specifications in infinite-horizon design of linear controllers) *are equivalent to explicit convex constraints on the parameters* H *of the underlying POB control law and therefore can be processed efficiently via Convex Optimization.*

Example: Discrete time H_∞ control. Discrete time H_∞ design specifications impose constraints on the behavior of the transfer matrix along the unit circumference $s = \exp\{\imath\omega\}$, $0 \le \omega \le 2\pi$, that is, on the steady state response of the closed loop system to a disturbance in the form of a harmonic oscillation.[4]. A rather general form of these specifications is a system of constraints

$$\|Q_i(s) - M_i(s)\mathcal{H}_{xu}(s)N_i(s)\| \le \tau_i \;\; \forall (s = \exp\{\imath\omega\} : \omega \in \Delta_i), \tag{5.4.27}$$

where $Q_i(s)$, $M_i(s)$, $N_i(s)$ are given rational matrix-valued functions with no singularities on the unit circumference $\{s : |s| = 1\}$, $\Delta_i \subset [0, 2\pi]$ are given segments, and $\|\cdot\|$ is the standard matrix norm (the largest singular value).

We are about to demonstrate that constraints (5.4.27) can be represented by an explicit finite system of LMIs; as a result, specifications (5.4.27) can be efficiently processed numerically. Here is the derivation. Both "transfer functions" $\mathcal{H}_x(s)$, $\mathcal{H}_u(s)$ are of the form $q^{-1}(s)Q(s,H)$,

[4]The entries of $\mathcal{H}_x(s)$ and $\mathcal{H}_u(s)$, restricted onto the unit circumference $s = \exp\{\imath\omega\}$, have very transparent interpretation. Assume that the only nonzero entry in the disturbances is the j-th one, and it varies in time as a harmonic oscillation of unit amplitude and frequency ω. The steady-state behavior of i-th state then will be a harmonic oscillation of the same frequency, but with another amplitude, namely, $|(\mathcal{H}_x(\exp\{\imath\omega))_{ij}|$ and phase shifted by $\arg((\mathcal{H}_x(\exp\{\imath\omega\})_{ij})$. Thus, the *state-to-input frequency responses* $(\mathcal{H}_x(\exp\{\imath\omega\}))_{ij}$ explain the steady-state behavior of states when the input is comprised of harmonic oscillations. The interpretation of the *control-to-input frequency responses* $(\mathcal{H}_u(\exp\{\imath\omega\}))_{ij}$ is completely similar.

where $q(s)$ is a scalar polynomial independent of H, and $Q(s,H)$ is a matrix-valued polynomial of s with coefficients *affinely depending on* H. With this in mind, we see that the constraints are of the generic form

$$\|p^{-1}(s)P(s,H)\| \le \tau\,\forall(s = \exp\{\imath\omega\} : \omega \in \Delta), \tag{5.4.28}$$

where $p(\cdot)$ is a scalar polynomial independent of H and $P(s,H)$ is a polynomial in s with $m \times n$ matrix coefficients affinely depending on H. Constraint (5.4.28) can be expressed equivalently by the semi-infinite matrix inequality

$$\left[\begin{array}{cc} \tau I_m & P(z,H)/p(z) \\ (P(z,H))^*/(p(z))^* & \tau I_n \end{array}\right] \succeq 0\,\forall(z = \exp\{\imath\omega\} : \omega \in \Delta)$$

(* stands for the Hermitian conjugate, $\Delta \subset [0, 2\pi]$ is a segment) or, which is the same,

$$\begin{array}{rcl} S_{H,\tau}(\omega) & \equiv & \left[\begin{array}{cc} \tau p(\exp\{\imath\omega\})(p(\exp\{\imath\omega\}))^* I_m & (p(\exp\{\imath\omega\}))^* P(\exp\{\imath\omega\}, H) \\ p(\exp\{\imath\omega\})(P(\exp\{\imath\omega\}, H))^* & \tau p(\exp\{\imath\omega\})(p(\exp\{\imath\omega\}))^* I_n \end{array}\right] \\ & \succeq & 0\,\forall\omega \in \Delta. \end{array}$$

Observe that $S_{H,\tau}(\omega)$ is a trigonometric polynomial taking values in the space of Hermitian matrices of appropriate size, the coefficients of the polynomial being affine in H, τ. It is known [49] that the cone $\mathcal{P}_m$ of (coefficients of) all Hermitian matrix-valued trigonometric polynomials $S(\omega)$ of degree $\le m$, which are $\succeq 0$ for all $\omega \in \Delta$, is semidefinite representable, i.e., there exists an explicit LMI

$$\mathcal{A}(S,u) \succeq 0$$

in variables S (the coefficients of a polynomial $S(\cdot)$) and additional variables u such that $S(\cdot) \in \mathcal{P}_m$ if and only if S can be extended by appropriate u to a solution of the LMI. Consequently, the relation

$$\mathcal{A}(S_{H,\tau}, u) \succeq 0, \tag{$*$}$$

which is an LMI in H, τ, u, is a semidefinite representation of (5.4.28): H, τ solve (5.4.28) if and only if there exists u such that H, τ, u solve $(*)$.

5.4.4 Putting Things Together: Infinite- and Finite-Horizon Design Specifications

For the time being, we have considered optimization over purified-output-based affine control laws in two different settings, finite- and infinite-horizon design specifications. In fact we can to some extent combine both settings, thus seeking affine purified-output-based controls ensuring both a good steady-state behavior of the closed loop system and a "good transition" to this steady-state behavior. The proposed methodology will become clear from the example that follows.

Consider the open-loop time-invariant system representing the discretized double-pendulum depicted on figure 5.2. The dynamics of the continuous time prototype plant is given by

$$\begin{array}{rcl} \dot{x} & = & A_c x + B_c u + R_c d \\ y & = & Cx, \end{array}$$

where

$$A_c = \left[\begin{array}{cccc} 0 & 1 & 0 & 0 \\ -1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & -1 & 0 \end{array}\right],\ B_c = \left[\begin{array}{c} 0 \\ 1 \\ 0 \\ 0 \end{array}\right],\ R_c = \left[\begin{array}{c} 0 \\ 0 \\ 0 \\ -1 \end{array}\right],\ C = \left[\begin{array}{cccc} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{array}\right]$$

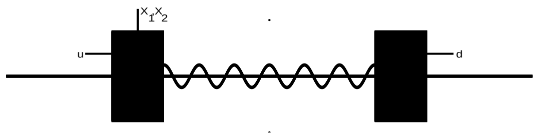


Figure 5.2: Double pendulum: two masses linked by a spring sliding without friction along a rod. Position and velocity of the first mass are observed.

(x_1, x_2 are the position and the velocity of the first mass, and x_3, x_4 those of the second mass). The discrete time plant we will actually work with is

$$\begin{array}{rcl} x_{t+1} & = & A_0 x_t + Bu_t + Rd_t \\ y_t & = & Cx_t \end{array} \tag{5.4.29}$$

where $A_0 = \exp\{\Delta \cdot A_c\}$, $B = \int_0^\Delta \exp\{sA_c\}B_c ds$, $R = \int_0^\Delta \exp\{sA_c\}R_c ds$. System (5.4.29) is not stable (absolute values of all eigenvalues of A_0 are equal to 1), which seemingly prevents us from addressing infinite-horizon design specifications via the techniques developed in section 5.4.3. The simplest way to circumvent the difficulty is to augment the original plant by a stabilizing time-invariant linear feedback; upon success, we then apply the purified-output-based synthesis to the augmented, already stable, plant. Specifically, let us look for a controller of the form

$$u_t = Ky_t + w_t. \tag{5.4.30}$$

With such a controller, (5.4.29) becomes

$$\begin{array}{rcl} x_{t+1} & = & Ax_t + Bw_t + Rd_t, \quad A = A_0 + BKC \\ y_t & = & Cx_t. \end{array} \tag{5.4.31}$$

If K is chosen in such a way that the matrix $A = A_0 + BKC$ is stable, we can apply all our purified-output-based machinery to the plant (5.4.31), with w_t in the role of u_t, however keeping in mind that the "true" controls u_t will be $Ky_t + w_t$.

For our toy plant, a stabilizing feedback K can be found by "brute force" — by generating a random sample of matrices of the required size and selecting from this sample a matrix, if any, which indeed makes (5.4.31) stable. Our search yielded feedback matrix $K = [-0.6950, -1.7831]$, with the spectral radius of the matrix $A = A_0 + BKC$ equal to 0.87. From now on, we focus on the resulting plant (5.4.31), which we intend to "close" by a control law from $\mathcal{C}_{8,0}$, where $\mathcal{C}_{k,0}$ is the family of all time invariant control laws of the form

$$w_t = \sum_{\tau=0}^{t} H_{t-\tau} v_\tau \quad \left[\begin{array}{rcl} v_t & = & y_t - C\widehat{x}_t, \\ \widehat{x}_{t+1} & = & A\widehat{x}_t + Bw_t, \ \widehat{x}_0 = 0 \end{array} \right] \tag{5.4.32}$$

where $H_s = 0$ when $s \geq k$. Our goal is to pick in $\mathcal{C}_{8,0}$ a control law with desired properties (to be precisely specified below) expressed in terms of the following 6 criteria:

- the four peak to peak gains z2x, z2u, d2x, d2u defined on p. 248;
- the two H_∞ gains

$$H_{\infty,x} = \max_{|s|=1,i,j} |(\mathcal{H}_x(s))|_{ij}, \quad H_{\infty,u} = \max_{|s|=1,i,j} |(\mathcal{H}_u(s))|_{ij},$$

Optimized criterion	Resulting values of the criteria					
	$z2x^{40}$	$z2u^{40}$	$d2x^{40}$	$d2u^{40}$	$H_{\infty,x}$	$H_{\infty,u}$
$z2x^{40}$	25.8	205.8	1.90	3.75	10.52	5.87
$z2u^{40}$	58.90	161.3	1.90	3.74	39.87	20.50
$d2x^{40}$	5773.1	13718.2	1.77	6.83	1.72	4.60
$d2u^{40}$	1211.1	4903.7	1.90	2.46	66.86	33.67
$H_{\infty,x}$	121.1	501.6	1.90	5.21	1.64	5.14
$H_{\infty,u}$	112.8	460.4	1.90	4.14	8.13	1.48
	z2x	z2u	d2x	d2u	$H_{\infty,x}$	$H_{\infty,u}$
(5.4.34)	31.59	197.75	1.91	4.09	1.82	2.04
(5.4.35)	2.58	0.90	1.91	4.17	1.77	1.63

Table 5.2: Gains for time invariant control laws of order 8 yielded by optimizing, one at a time, the criteria $z2x^{40}$,...,$H_{\infty,u}$ over control laws from $\mathcal{F} = \{\eta \in \mathcal{C}_{8,0} : d2x^{40}[\eta] \leq 1.90\}$ (first six lines), and by solving programs (5.4.34), (5.4.35) (last two lines).

where $\mathcal{H}_x$ and $\mathcal{H}_u$ are the transfer functions from the disturbances to the states and the controls, respectively.

Note that while the purified-output-based control w_t we are seeking is defined in terms of the stabilized plant (5.4.31), the criteria z2u,d2u, $H_{\infty,u}$ are defined in terms of the original controls $u_t = Ky_t + w_t = KCx_t + w_t$ affecting the actual plant (5.4.29).

In the synthesis we are about to describe our primary goal is to minimize the global disturbance to state gain d2x, while the secondary goal is to avoid too large values of the remaining criteria. We achieve this goal as follows.

Step 1: Optimizing d2x**.** As it was explained on p. 248, the optimization problem

$$\mathrm{Opt}_{\mathrm{d2x}}(k,0;N_+) = \min_{\eta\in\mathcal{C}_{k,0}} \max_{0\leq t\leq N_+} \mathrm{d2x}_t[\eta] \tag{5.4.33}$$

is an explicit convex program (in fact, just an LO), and its optimal value is a lower bound on the best possible global gain d2x achievable with control laws from $\mathcal{C}_{k,0}$. In our experiment, we solve (5.4.33) for $k = 8$ and $N_+ = 40$, arriving at $\mathrm{Opt}_{\mathrm{d2x}}(8,0;40) = 1.773$. The global d2x gain of the resulting time-invariant control law is 1.836 — just 3.5% larger than the outlined lower bound. We conclude that the control yielded by the solution to (5.4.33) is nearly the best one, in terms of the global d2x gain, among time-invariant controls of order 8. At the same time, part of the other gains associated with this control are far from being good, see line "$d2x^{40}$" in table 5.2.

Step 2: Improving the remaining gains. To improve the "bad" gains yielded by the nearly d2x-optimal control law we have built, we act as follows: we look at the family $\mathcal{F}$ of all time invariant control laws of order 8 with the finite-horizon d2x gain $\mathrm{d2x}^{40}[\eta] = \max\limits_{0\leq t\leq 40} \mathrm{d2x}_t[\eta]$ not exceeding 1.90 (that is, look at the controls from $\mathcal{C}_{8,0}$ that are within 7.1% of the optimum in terms of their $d2x^{40}$ gain) and act as follows:

A. We optimize over $\mathcal{F}$, one at a time, every one of the remaining criteria $\mathrm{z2x}^{40}[\eta] = \max\limits_{0\leq t\leq 40} \mathrm{z2x}_t[\eta]$, $\mathrm{z2u}^{40}[\eta] = \max\limits_{0\leq t\leq 40} \mathrm{z2u}_t[\eta]$, $\mathrm{d2u}^{40}[\eta] = \max\limits_{0\leq t\leq 40} \mathrm{d2u}_t[\eta]$, $H_{\infty,x}[\eta]$, $H_{\infty,u}[\eta]$, thus obtaining "reference values" of these criteria; these are lower bounds on the optimal values of the corresponding global gains, optimization being carried out over the set $\mathcal{F}$. These lower bounds are the underlined data in table 5.2.

B. We then minimize over $\mathcal{F}$ the "aggregated gain"

$$\frac{\mathrm{z2x}^{40}[\eta]}{25.8} + \frac{\mathrm{z2u}^{40}[\eta]}{161.3} + \frac{\mathrm{d2u}^{40}[\eta]}{2.46} + \frac{H_{\infty,x}[\eta]}{1.64} + \frac{H_{\infty,u}[\eta]}{1.48} \tag{5.4.34}$$

(the denominators are exactly the aforementioned reference values of the corresponding gains). The global gains of the resulting time-invariant control law of order 8 are presented in the "(5.4.34)" line of table 5.2.

Step 3: Finite-horizon adjustments. Our last step is to improve the z2x and z2u gains by passing from a time invariant affine control law of order 8 to a nearly time invariant law of order 8 with stabilization time $N_* = 20$. To this end, we solve the convex optimization problem

$$\min_{\eta\in\mathcal{C}_{8,20}} \left\{ \mathrm{z2x}^{50}[\eta] + \mathrm{z2u}^{50}[\eta] : \begin{array}{rcl} \mathrm{d2x}^{50}[\eta] & \le & 1.90 \\ \mathrm{d2u}^{50}[\eta] & \le & 4.20 \\ H_{\infty,x}[\eta] & \le & 1.87 \\ H_{\infty,u}[\eta] & \le & 2.09 \end{array} \right\} \tag{5.4.35}$$

(the right hand sides in the constraints for $\mathrm{d2u}^{50}[\cdot]$, $H_{\infty,x}[\cdot]$, $H_{\infty,u}[\cdot]$ are the slightly increased (by 2.5%) gains of the time invariant control law obtained in Step 2). The global gains of the resulting control law are presented in the last line of table 5.2, see also figure 5.3. We see that finite-horizon adjustments allow us to reduce by orders of magnitude the global z2x and z2u gains and, as an additional bonus, result in a substantial reduction of H_∞-gains.

Simple as this control problem may be, it serves well to demonstrate the importance of purified-output-based representation of affine control laws and the associated possibility to express various control specifications as explicit convex constraints on the parameters of such laws.

5.5 Exercises

Exercise 5.1 Consider a discrete time linear dynamical system

$$\begin{array}{rcl} x_0 & = & z \\ x_{t+1} & = & A_t x_t + B_t u_t + R_t d_t,\ t = 0, 1, ... \end{array} \tag{5.5.1}$$

where $x_t \in \mathbb{R}^n$ are the states, $u_t \in \mathbb{R}^m$ are the controls, and $d_t \in \mathbb{R}^k$ are the exogenous disturbances. We are interested in the behavior of the system on the finite time horizon $t = 0, 1, ..., N$. A "desired behavior" is given by the requirement

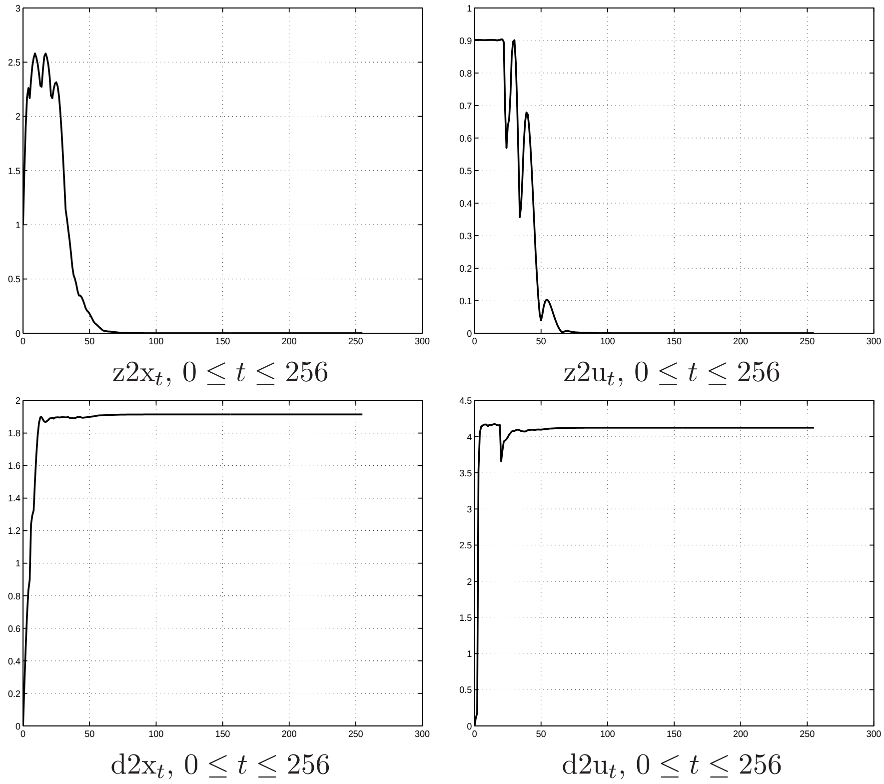
$$\|Pw^N - q\|_\infty \le R \tag{5.5.2}$$

on the state-control trajectory $w^N = [x_0; ...; x_{N+1}; u_0; ...; u_N]$.

Let us treat $\zeta = [z; d_0; ...; d_N]$ as an uncertain perturbation with perturbation structure $(\mathcal{Z}, \mathcal{L}, \|\cdot\|_r)$, where

$$\mathcal{Z} = \{\zeta : \|\zeta - \bar{\zeta}\|_s \le R\},\ \mathcal{L} = \mathbb{R}^L \qquad [L = \dim \zeta]$$

and $r, s \in [1, \infty]$, so that (5.5.1), (5.5.2) become a system of uncertainty-affected linear constraints on w^N. We want to process the Affinely Adjustable GRC of the system, where u_t are allowed to be affine functions of the initial state z and the vector of disturbances $d^t = [d_0; ...; d_t]$ up to time t, and the states x_t are allowed to be affine functions of z and d^{t-1}. We wish to minimize the corresponding global sensitivity.



z2x$_t$, $0 \leq t \leq 256$

z2u$_t$, $0 \leq t \leq 256$

d2x$_t$, $0 \leq t \leq 256$

d2u$_t$, $0 \leq t \leq 256$

Figure 5.3: Frequency responses and gains of control law given by solution to (5.4.35).

In control terms: we want to "close" the open-loop system (5.5.1) with a non-anticipative affine control law

$$u_t = U_t^z z + U_t^d d^t + u_t^0 \tag{5.5.3}$$

based on observations of initial states and disturbances up to time t in such a way that the "closed loop system" (5.5.1), (5.5.3) exhibits the desired behavior in a robust w.r.t. the initial state and the disturbances fashion.

Write down the AAGRC of our uncertain problem as an explicit convex program with efficiently computable constraints.

Exercise 5.2 Consider the modification of Exercise 5.1 where the cone $\mathcal{L} = \mathbb{R}^L$ is replaced with

$$\mathcal{L} = \{[0; d_0; ...; d_N] : d_t \geq 0,\, 0 \leq t \leq N\},$$

and solve the corresponding version of the Exercise.

Exercise 5.3 Consider the simplest version of Exercise 5.1, where (5.5.1) reads

$$\begin{array}{rcl} x_0 &=& z \in \mathbb{R} \\ x_{t+1} &=& x_t + u_t - d_t,\, t = 0, 1, ..., 15, \end{array}$$

(5.5.2) reads

$$|\theta x_t| = 0,\, t = 1, 2, ..., 16,\;\; |u_t| = 0,\, t = 0, 1, ..., 15$$

and the perturbation structure is

$$\mathcal{Z} = \{[z; d_0; ...; d_{15}] = 0\} \subset \mathbb{R}^{17},\, \mathcal{L} = \{[0; d_0; d_1; ...; d_{15}]\}, \|\zeta\| \equiv \|\zeta\|_2.$$

Assuming the same "adjustability status" of u_t and x_t as in Exercise 5.1,

1. Represent the AAGRC of (the outlined specializations of) (5.5.1), (5.5.2), where the goal is to minimize the global sensitivity, as an explicit convex program;

2. Interpret the AAGRC in Control terms;

3. Solve the AAGRC for the values of θ equal to 1.e6, 10, 2, 1.

Exercise 5.4 Consider a communication network — an oriented graph G with the set of nodes $V = \{1, ..., n\}$ and the set of arcs Γ. Several ordered pairs of nodes (i, j) are marked as "source-sink" nodes and are assigned traffic d_{ij} — the amount of information to be transmitted from node i to node j per unit time; the set of all source-sink pairs is denoted by $\mathcal{J}$. Arcs $\gamma \in \Gamma$ of a communication network are assigned with capacities — upper bounds on the total amount of information that can be sent through the arc per unit time. We assume that the arcs already possess certain capacities p_γ, which can be further increased; the cost of a unit increase of the capacity of arc γ is a given constant c_γ.

1) Assuming the demands d_{ij} certain, formulate the problem of finding the cheapest extension of the existing network capable to ensure the required source-sink traffic as an LO program.

2) Now assume that the vector of traffic $d = \{d_{ij} : (i, j) \in \mathcal{J}\}$ is uncertain and is known to run through a given semidefinite representable compact uncertainty set $\mathcal{Z}$. Allowing the amounts x_γ^{ij} of information with origin i and destination j traveling through the arc γ to depend affinely on traffic, build the AARC of the (uncertain version of the) problem from 1). Consider

two cases: (a) for every $(i,j) \in \mathcal{J}$, x_γ^{ij} can depend affinely solely on d_{ij}, and (b) x_γ^{ij} can depend affinely on the entire vector d. Are the resulting problems computationally tractable?

3) Assume that the vector d is random, and its components are independent random variables uniformly distributed in given segments Δ_{ij} of positive lengths. Build the chance constrained versions of the problems from 2).