

A randomized relaxation method to ensure feasibility in stochastic control of linear systems subject to state and input constraints

Luca Deori^a Simone Garatti^a Maria Prandini^a

^aDipartimento di Elettronica, Informazione e Bioingegneria, Politecnico di Milano, Via Ponzio 34/5, 20133 Milano, Italy

Abstract

We consider a linear system affected by an additive stochastic disturbance and address the design of a finite horizon control policy that is optimal according to some cost criterion and accounts also for probabilistic constraints on both the input and state variables. The resulting policy can be implemented over a receding horizon according to the model predictive control strategy. Such a possibility, however, is hampered by the fact that a feasibility issue may arise when recomputing the policy. Infeasibility indeed can occur if the disturbance has unbounded support and the state is required to remain in a bounded set. In this paper, we propose a solution to this issue that is based on the introduction of a constraint relaxation that becomes effective only when the original problem turns out to be unfeasible. This is obtained via a cascade of two probabilistically-constrained optimization problems where, in the first one, performance is neglected and the policy is designed to fully recover feasibility or –if this is not possible– to determine the minimum level of relaxation which is needed to recover feasibility; in the second step, such a minimum relaxation level is imposed while optimally (re-)tuning the control policy parameters. Both problems are solved through a computationally tractable scenario-based scheme using a finite number of disturbance realizations and providing an approximate solution that satisfies with high confidence the original probabilistic constraints of the cascade.

Key words: stochastic constrained control, randomized methods, scenario approach, model predictive control

1 Problem formulation

This paper deals with the problem of designing a finite-horizon optimal control policy for a discrete-time stochastic linear system subject to constraints on both the state and the input. Specifically, we consider the system

$$x_{t+1} = Ax_t + Bu_t + B_w w_t, \quad (1)$$

where $x_t \in \mathbb{R}^n$ is the state, $u_t \in \mathbb{R}^m$ is the control input and $w_t \in \mathbb{R}^{n_w}$ is an additive stochastic disturbance. Matrices A , B , and B_w have appropriate dimensions so as to make (1) consistent. The probability distribution of w_t is assumed to be known and it may have an unbounded support. Without loss of generality, we assume that $n_w \leq n$ and B_w is full column rank. The state is accessible, i.e., at every t a noise-free measurement of x_t becomes available.

The following disturbance feedback parametrization for the

control input is adopted:

$$u_t = \begin{cases} \gamma_0, & t = 0 \\ \gamma + \sum_{\tau=0}^{t-1} \theta_{t,\tau} w_\tau, & t > 0 \end{cases} \quad (2)$$

where $\gamma \in \mathbb{R}^m$ represent open-loop terms, while $\theta_{t,\tau} \in \mathbb{R}^{m \times n_w}$ are the disturbance feedback gains. Note that the stochastic disturbance w_τ , $\tau = 0, \dots, t-1$, appearing in (2) to determine the input at time $t > 0$ can be recovered from the measurements of the state x_τ , $\tau = 0, \dots, t$, according to

$$w_\tau = B_w^\dagger (x_{\tau+1} - Ax_\tau - Bu_\tau), \quad (3)$$

where B_w^\dagger denotes the pseudo-inverse of B_w . This expression reveals that the disturbance feedback control policy in (2) is in fact a state feedback control policy. Parametrization (2) was first proposed in [24], where it was shown that the family of policies in (2) is indeed equivalent to the family of affine state feedback policies $u_t = \tilde{\gamma}_t + \sum_{\tau=0}^t \tilde{\theta}_{t,\tau} x_\tau$.¹ To be precise, for every choice of $\tilde{\gamma}_t, \tilde{\theta}_{t,\tau}$ there exists a parametrization $\gamma, \theta_{t,\tau}$ in (2) returning the same control action, and viceversa.

* Research was supported by the European Commission under the project UnCoVerCPS, grant number 643921.

**Corresponding author S. Garatti. Ph. +39-02-23993650. Fax +39-02-23993412.

Email addresses: luca.deori@polimi.it (Luca Deori), simone.garatti@polimi.it (Simone Garatti), maria.prandini@polimi.it (Maria Prandini).

¹ By setting $\tilde{\theta}_{t,\tau} = 0$, $\tau \neq t$, and $\tilde{\theta}_{t,t} = K$, this expression simplifies to $u_t = Kx_t + \tilde{\gamma}_t$, which can then be represented via (2), thanks to the equivalence result in [24].

The great advantage of (2) is that, differently from other parameterizations, the input u_t and the state x_t are affine functions of the design parameters γ_t and $\theta_{t,\tau}$, which yields clear computational benefits and allows one to tune the state feedback gain. This is not so for other parameterizations and indeed, in the literature where the policy $u_t = Kx_t + c_t$ is adopted, it is often the case that the gain K is fixed to an a-priori chosen value to recover linearity, [15].

Remark 1 (feedforward compensator) *Note that, if the disturbance were directly measured, we would be addressing the design of a feedforward disturbance compensator. In such a setting, u_t in (2) could depend also on w_t and the results in the paper would still apply with minor modifications.* ■

The objective is to design the parameters γ_t and $\theta_{t,\tau}$ in (2) so as to minimize a cost function over a finite time horizon of length M , while accounting for constraints on the input and state variables. This problem may arise per-se in some applications (for instance, the positioning of the end-effector of an industrial robot equipped with a robot re-initialization device), but its significance mainly lies in the fact that it occurs in Model Predictive Control (MPC), where the problem is repeatedly solved over a receding horizon, [29,9,35,27]. In our formulation, we admit as cost any strictly convex function J of the parameters γ_t and $\theta_{t,\tau}$ over the horizon $0, 1, \dots, M-1$. Plainly, the most common situation is when J is defined as a function of the input and the state. A typical choice is the average quadratic cost

$$J = \mathbb{E} \left[\sum_{t=1}^M x_t^T Q x_t + \sum_{t=0}^{M-1} u_t^T R u_t \right], \quad (4)$$

where Q and R are symmetric and positive semi-definite matrices, and \mathbb{E} denotes expectation with respect to the (known) distribution of the disturbance. In this case, a sufficient condition for strict convexity to hold is that matrices R and $\mathbb{E}[\mathbf{w}\mathbf{w}^T]$ are positive definite, see [21].

As for the input and state constraints, we assume that they are expressed as

$$f(u_0, \dots, u_{M-1}) \leq 0 \wedge g(x_1, \dots, x_M, u_0, \dots, u_{M-1}) \leq 0 \quad (5)$$

where \wedge stands for “and”, $f: \mathbb{R}^{mM} \rightarrow \mathbb{R}^{p_u}$ and $g: \mathbb{R}^{(n+m)M} \rightarrow \mathbb{R}^{p_y}$ are continuous convex vector-valued functions, and the inequalities are meant component-wise. For example, a typical requirement is that the norms of the input and of some output variable are kept within an admissible range for each t in the time horizon. Note that g allows for joint state and input constraints and the constraints expressed by f could be incorporated in g . To ease further explanations, we however keep the constraints that depend on the input only separate from the others.

It should be noted that constraints (5) cannot be directly imposed since they miss to specify how to account for the presence of the stochastic disturbance affecting both the state

and the input variables. Robustly enforcing the state constraint for every and each disturbance realization is impossible when the state is constrained within a bounded set and the disturbance has unbounded support. As a matter of fact, the stochastic disturbance w_t enters additively the system dynamics, and, since the input u_t depends on the disturbance up to time $t-1$ at most, the dependence of x_{t+1} on w_t cannot be canceled. We then assume that constraints are enforced probabilistically, namely, constraints (5) are required to hold with a certain (usually high) probability $1 - \varepsilon$, where $\varepsilon \in (0, 1)$ is a user-chosen parameter:

$$\mathbb{P}\{f(u_0, \dots, u_{M-1}) \leq 0 \wedge g(x_1, \dots, x_M, u_0, \dots, u_{M-1}) \leq 0\} \geq 1 - \varepsilon. \quad (6)$$

This probabilistic formulation of constraints is the most natural for many problems of interest and has become common in the recent literature on constrained stochastic control, [3,31,33,34,4,18,15,22,23,28].

Altogether, the optimal design problem we are considering is as follows:

$$\min_{\gamma_t, \theta_{t,j}} J \text{ subject to (6)}. \quad (7)$$

Note that a probabilistic constraint $\mathbb{P}\{f(u_0, \dots, u_{M-1}) \leq 0\} \geq 1 - \varepsilon$, where g is not present, is always feasible, because, if needed, the disturbance feedback gains $\theta_{t,\tau} \in \mathbb{R}^{m \times n_w}$ in (2) can be set to zero, which makes u_t deterministic. On the contrary, a feasibility issue arises in (7) precisely because of the presence of the requirement on $g(x_1, \dots, x_M, u_0, \dots, u_{M-1}) \leq 0$. Since the additive disturbance w_t has possibly unbounded support and given the limitation imposed by the system dynamics and by the constraints on the input variable, it may then be that, depending on the system initialization x_0 , no choice of $\gamma_t, \theta_{t,\tau}$ exists such that $g(x_1, \dots, x_M, u_0, \dots, u_{M-1}) \leq 0$ is attained with the required probabilistic level. Just to give some intuition, this is typically the case when the state is constrained within a bounded set and x_0 is close to the boundary: the values of w_0 that make x_1 exit the constraint set can have a probability larger than $1 - \varepsilon$, irrespectively of the value chosen for u_0 within the assigned actuation limits. The feasibility issue gets worse for large values of M . For instance, when the noise is Gaussian and B_w is the identity matrix, no matter how u is chosen, the state will exit any fixed bounded set with probability 1 as the time horizon length grows unbounded, [26]. The feasibility problem here discussed is severe because in many cases the designer has no direct control on the system initialization, which is indeed determined by exogenous causes. For example, in an MPC scheme where the optimization problem (7) is continuously repeated at every time step over a receding horizon, the system initialization for a given time horizon is determined by the solutions at previous steps. Since constraints are only probabilistically enforced, it may well be that at previous steps the state is driven to a value preventing feasibility at the current iteration.

The objective of this paper is that of addressing the feasibility issue illustrated above. To this purpose we introduce a suitable relaxation of problem (7), which is inspired by constraint softening in deterministic optimization and is conceived so as to adhere to the intent of the original problem formulation (7) as much as possible. Precisely, whenever the original constraint (6) is feasible, the original problem is maintained, while, otherwise, a new decision problem is formulated by relaxing the condition $g(x_1, \dots, x_M, u_0, \dots, u_{M-1}) \leq 0$ only for those components of the vector inequality that need to be relaxed to get feasibility. Constraint relaxation is set to a minimal level needed to recover feasibility so as to avoid penalizing excessively performance. In particular, no relaxation is introduced if the problem is feasible.

This reformulation leads to a cascade of two optimization problems with probabilistic constraints, which, admittedly, can be very difficult to solve in general, since problems involving probabilistic constraints can be NP-hard. The second contribution of this paper is that of introducing a resolution scheme based on randomization in order to enhance computational tractability. Specifically, we resort to the so-called scenario approach, [5,6,11,14], a recently introduced randomized method that can be used to provide approximate solutions to problems with probabilistic constraints establishing a precise link between the original problem and its approximation. In this paper, such a link is extended to the scenario solution to the cascade of problems discussed above, which is a non-standard setup not fully covered by the available literature (see [30] for a contribution on cascading optimization in a different setup).

1.1 Literature review

In [15,17] stochastic uncertainty with bounded support is tackled by means of suitable probabilistic tubes, whereas in [28] constraint tightening is adopted to enforce recursive feasibility in MPC, always under the assumption of a bounded disturbance. In the case of systems affected by stochastic disturbance with *unbounded support*, control problems in presence of state constraints have been addressed in [2,25,3,31,33,34,4,18]. In [2,33,34], state constraints are dealt with by means of a penalization term accounting for the state constraint violation so as to ensure feasibility. In [25,31,4,18], an analytic convex relaxation of probabilistic constraints is proposed, whereas in [3] the problem is reformulated considering a bounded disturbance obtained by suitably cutting the tails of the disturbance distribution. In all these approaches, the disturbance is assumed to be a sequence of i.i.d. (independent and identically distributed) random variables. Many of them also assume that the disturbance has a Gaussian distribution, [2,25,3,31,4,18]. This paper addresses the unbounded disturbance case and differs from the mentioned approaches in that no independence and Gaussianity assumptions are made.

A main feature of our method is that a randomized approach is adopted to recover tractability. Other randomized approaches to constrained stochastic control for system (1)

have been also proposed in [7,36,8] but under the assumption that the noise has bounded support, whereas in [32] only input constraints are considered, and in [37,40] recursive feasibility is assumed.

Our previous contributions [19,21] address the same set-up as in this paper but they recover feasibility by either adding a term penalizing state constraint violation to the cost or introducing a certain pre-defined admissible deterioration of the system performance while relaxing the state constraints. Choosing the appropriate weight of the penalization term or degree of performance deterioration requires some tuning effort. Also, it might be the case that constraint relaxation is introduced even if the randomized problem is feasible, [19,21], which is not the case in our method.

A preliminary version of this paper has been presented as a conference contribution in [20], where the effectiveness of the proposed randomized solution to the feasibility issue was shown via a numerical case study only. The current submission provides revised and new discussions and a theoretical analysis with formal proof of our main technical achievement.

1.2 Structure of the paper

The rest of the paper is organized as follows. Some compact notation is introduced in Section 2. In Section 3, we formally introduce the problem relaxation, while the proposed algorithmic resolution scheme based on the scenario approach is described in Section 4. In this section, the theoretical properties of the obtained solution are also discussed and a formal proof of these properties is reported. A numerical example is given in Section 5, while some concluding remarks are drawn in Section 6.

2 Compact notation

In order to ease the notation we define:

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_M \end{bmatrix} \quad \mathbf{u} = \begin{bmatrix} u_0 \\ u_1 \\ \vdots \\ u_{M-1} \end{bmatrix} \quad \mathbf{w} = \begin{bmatrix} w_0 \\ w_1 \\ \vdots \\ w_{M-1} \end{bmatrix} \quad \mathbf{F} = \begin{bmatrix} A \\ A^2 \\ \vdots \\ A^M \end{bmatrix}$$

$$\mathbf{G} = \begin{bmatrix} B & 0_{n \times m} & \cdots & 0_{n \times m} \\ AB & B & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0_{n \times m} \\ A^{M-1}B & \cdots & AB & B \end{bmatrix} \quad \mathbf{H} = \begin{bmatrix} B_w & 0_{n \times n_w} & \cdots & 0_{n \times n_w} \\ AB_w & B_w & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0_{n \times n_w} \\ A^{M-1}B_w & \cdots & AB_w & B_w \end{bmatrix},$$

so that the state vector can be calculated as:

$$\mathbf{x} = \mathbf{F}x_0 + \mathbf{G}\mathbf{u} + \mathbf{H}\mathbf{w}. \quad (8)$$

Similarly, the disturbance feedback policy (2) can be rewritten in the following compact form

$$\mathbf{u} = \Gamma + \Theta \mathbf{w}, \quad (9)$$

where we let

$$\Gamma = \begin{bmatrix} \gamma_0 \\ \gamma_1 \\ \vdots \\ \gamma_{M-1} \end{bmatrix} \quad \Theta = \begin{bmatrix} 0_{m \times n_w} & 0_{m \times n_w} & \cdots & 0_{m \times n_w} \\ \theta_{1,0} & 0_{m \times n_w} & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0_{m \times n_w} \\ \theta_{M-1,0} & \cdots & \theta_{M-1,M-2} & 0_{m \times n_w} \end{bmatrix}.$$

By substituting the expression of the input in (9) into (8), the affine dependence of \mathbf{x} on the design parameters Γ and Θ becomes clear:

$$\mathbf{x} = \mathbf{F}\mathbf{x}_0 + \mathbf{G}\Gamma + (\mathbf{G}\Theta + \mathbf{H})\mathbf{w}$$

Eventually, the nonzero components of Γ and Θ are collected in the vector of optimization variables λ , so that the following notations can be adopted: $\mathbf{u} = \mathbf{u}_\lambda(\mathbf{w})$, $\mathbf{x} = \mathbf{x}_\lambda(\mathbf{w})$, and $J = J(\lambda)$, which point out the dependence of input, state, and cost on the optimization vector λ and the disturbance realization \mathbf{w} . The constraints in (6) then become

$$\mathbb{P}\{f(\mathbf{u}_\lambda(\mathbf{w})) \leq 0 \wedge g(\mathbf{x}_\lambda(\mathbf{w}), \mathbf{u}_\lambda(\mathbf{w})) \leq 0\} \geq 1 - \varepsilon.$$

3 Problem relaxation to ensure feasibility

In order to recover feasibility, we introduce a relaxation of the condition $g(\mathbf{x}_\lambda(\mathbf{w}), \mathbf{u}_\lambda(\mathbf{w})) \leq 0$ by substituting its right-hand side with $h \in \mathbb{R}^{p_y}$, h being a new optimization variable. By doing this, the constraint involving state variables turns out to be always feasible, since it is enough to take the variable h large enough. On the other hand, large values for h are clearly not desired since the bigger h the larger the deviation from the original constraint. To stick to the original problem formulation as much as possible h should be minimized component-wise. On the other hand, one should account for the minimization of the cost function $J(\lambda)$, which represents the system performance. This recalls the constraint softening approach adopted in deterministic optimization to cope with constraint infeasibility. In classical constraint softening, a penalized cost function is minimized to compromise between constraint softening and performance optimization. Here, in order to drive the system as close as possible to the feasibility condition, we first compute the minimum value for the parameter h to make the relaxed constraint feasible, without accounting for performance, and optimize performance only afterwards, while enforcing the relaxed constraint with the computed minimum value for h . More precisely, the following cascade of optimization programs (two-step approach) is proposed, where $L(h)$ is an

user-chosen strictly convex function of h that is also positive definite at $h = 0$ (i.e., $L(h) > 0$, $h \neq 0$ and $L(0) = 0$):

$$\min_{\lambda, h \geq 0} L(h) \text{ subject to:} \quad (10)$$

$$\mathbb{P}\{f(\mathbf{u}_\lambda(\mathbf{w})) \leq 0 \wedge g(\mathbf{x}_\lambda(\mathbf{w}), \mathbf{u}_\lambda(\mathbf{w})) \leq h\} \geq 1 - \varepsilon, \quad (11)$$

$$\min_{\lambda} J(\lambda) \text{ subject to:} \quad (11)$$

$$\mathbb{P}\{f(\mathbf{u}_\lambda(\mathbf{w})) \leq 0 \wedge g(\mathbf{x}_\lambda(\mathbf{w}), \mathbf{u}_\lambda(\mathbf{w})) \leq h^o\} \geq 1 - \varepsilon,$$

where h^o is the optimal value for h obtained in (10). The optimal value for λ obtained from (11) is denoted by λ^o . Problem (10) in the first step aims at determining the smallest value of h , according to the cost $L(h)$, that ensures the feasibility of the probabilistic constraint

$$\mathbb{P}\{f(\mathbf{u}_\lambda(\mathbf{w})) \leq 0 \wedge g(\mathbf{x}_\lambda(\mathbf{w}), \mathbf{u}_\lambda(\mathbf{w})) \leq h\} \geq 1 - \varepsilon.$$

A possible choice for the cost function $L(h)$ is e.g. $L(h) = h^T T h$, which allows to assign a different importance to each component of h by properly choosing the positive definite matrix T . Note that since the cost function $L(h)$ does not depend on λ , it may happen that the optimal cost $L(h^o)$ is attained for multiple λ values, each of them leading to a possibly different value of $J(\lambda)$. The second step optimization problem (11) then exploits this degree of freedom to minimize the performance cost. To this purpose, $J(\lambda)$ is minimized while the relaxed constraint $\mathbb{P}\{f(\mathbf{u}_\lambda(\mathbf{w})) \leq 0 \wedge g(\mathbf{x}_\lambda(\mathbf{w}), \mathbf{u}_\lambda(\mathbf{w})) \leq h^o\} \geq 1 - \varepsilon$ is enforced. Since the bound on the state condition $g(\mathbf{x}_\lambda(\mathbf{w}), \mathbf{u}_\lambda(\mathbf{w}))$ is fixed to h^o as computed in the previous step, problem (11) does not suffer from any feasibility issue. The cascade of problems is conceived so that, when the probabilistic constraint in (7) is infeasible, the control action is basically designed according to (10) so as to recover feasibility (minimization of $L(h)$). In this case, (11) provides just a refinement of the solution. The requirement $h \geq 0$ in (10) ensures that the constraint relaxation in (11), component by component, cannot become tighter than the original constraint in (7), and, for those components not requiring any relaxation, (11) pursues the goal of minimizing $J(\lambda)$ as in (7). In particular, whenever (7) is already feasible, program (10) simply returns $h^o = 0$ and the original problem (7) is recovered in (11).

Remark 2 Another relaxation approach, which is quite popular in optimization and which was studied in [19,21], consists in introducing the same relaxation adopted here by adding the optimization variable h , but solving a single optimization problem with $J(\lambda) + \mu L(h)$ as cost function to be minimized:

$$\min_{\lambda, h \geq 0} J(\lambda) + \mu L(h) \text{ subject to:} \quad (12)$$

$$\mathbb{P}\{f(\mathbf{u}_\lambda(\mathbf{w})) \leq 0 \wedge g(\mathbf{x}_\lambda(\mathbf{w}), \mathbf{u}_\lambda(\mathbf{w})) \leq h\} \geq 1 - \varepsilon,$$

The drawbacks of this approach are that an appropriate tuning of the penalization parameter $\mu > 0$ is necessary to get the desired compromise between constraint relaxation and

performance, see [19,21]. This tuning may be impractical and difficult to achieve, and it may happen that some relaxation is introduced even when it is not needed. A further more detailed discussion on these aspects is presented in Remark 3, in the light of some technical results, and in Section 5, where the approach of this paper is compared to that of [19,21] in a numerical example.

Overall, the cascade of problems (10) and (11) returns a solution given by the pair (λ^o, h^o) , where λ^o determines the control action to be implemented and h^o is the probabilistically guaranteed bound for the state constraint. Note that the value h^o computed in the first step optimization problem can be inspected to evaluate the mismatch with respect to the original state constraint.

4 Scenario-based resolution scheme

Problems (10) and (11) are, in general, hard to solve because of the presence of a probabilistic constraint. Hence, we propose to tackle them by means of a sample-based scheme, which is in the vein of the so-called scenario approach, [5,6,11,14]. The proposed scheme allows to recover computational tractability at the price of introducing some approximation. However, by exploiting the scenario approach theory, which is here extended to the cascade of problems (10) and (11), precise probabilistic guarantees on the feasibility of the achieved solution are also provided.

The idea of the scenario approach is to consider N disturbance realizations of length M :

$$\mathbf{w}^{(k)} = \begin{bmatrix} w_0^{(k)} & w_1^{(k)} & \dots & w_{M-1}^{(k)} \end{bmatrix}, k = 1, \dots, N,$$

each one extracted according to the disturbance probability distribution. Then, the probabilistic constraint in (10) and (11) are replaced by N non-probabilistic constraints, one for each disturbance realizations. More precisely, we have the following cascade of problems that can be seen as a sample-based counterpart of the cascade of problems (10) and (11):

$$\min_{\lambda, h \geq 0} L(h) \text{ subject to:} \quad (13)$$

$$f(\mathbf{u}_\lambda(\mathbf{w}^{(k)})) \leq 0, k = 1, \dots, N,$$

$$g(\mathbf{x}_\lambda(\mathbf{w}^{(k)}), \mathbf{u}_\lambda(\mathbf{w}^{(k)})) \leq h, k = 1, \dots, N,$$

$$\min_{\lambda} J(\lambda) \text{ subject to:} \quad (14)$$

$$f(\mathbf{u}_\lambda(\mathbf{w}^{(k)})) \leq 0, k = 1, \dots, N,$$

$$g(\mathbf{x}_\lambda(\mathbf{w}^{(k)}), \mathbf{u}_\lambda(\mathbf{w}^{(k)})) \leq h^*, k = 1, \dots, N,$$

where h^* is the optimal value of h obtained in (13). The optimal value for λ obtained from (14) is denote by λ^* . Problems (13) and (14) are convex and have a finite number of constraints; hence, they can be efficiently solved by means of standard solvers. Note that as the constraints are convex and the cost function $L(h)$ is strictly convex with respect to its argument, problem (13) uniquely determines the

value of h^* ; similarly, thanks to the strict convexity of $J(\lambda)$, the solution to problem (14) is unique.²

The same interpretation we had for the cascade of problems (10) and (11) in Section 3 applies to the cascade of problems (13) and (14): indeed, the solution of the latter cascade defined by the pair (λ^*, h^*) is the empirical counterpart of the solution of the former. It is worth noticing that, as the pair (λ^*, h^*) is feasible and optimal for (13), the second step optimization problem (14) can be regarded as a tie break rule by means of which the solution that minimizes the cost $J(\lambda)$ is chosen among the possible multiple solutions in λ of the first step optimization problem (13).

We are now interested in studying the feasibility of the obtained scenario-based solution for the probabilistic constraint

$$\mathbb{P} \{f(\mathbf{u}_\lambda(\mathbf{w})) \leq 0 \wedge g(\mathbf{x}_\lambda(\mathbf{w}), \mathbf{u}_\lambda(\mathbf{w})) \leq h\} \geq 1 - \varepsilon, \quad (15)$$

so as to provide a connection between (λ^*, h^*) and the original cascade of problems (10) and (11).

This question pertains to the theory of the scenario approach, which provides in a number of different setups guarantees on the feasibility of the scenario solution for the original probabilistic constraint as long as N is suitably chosen, see e.g. [6,11,12,10,16,13]. The tightest result is that of [11] which, however, does not directly apply to the cascade of problems (13) and (14). The results on cascading optimization in [30] apply to this context but the resulting bound on N is conservative. Theorem 1 provides an extension of the result in [11] to the current framework, under the following assumption.

Assumption 1 *The set of admissible control actions $\{(u_0, \dots, u_{M-1}) : f(u_0, \dots, u_{M-1}) \leq 0\}$ has nonempty interior.* ■

Note that Assumption 1 simply requires that the constraint $f(u_0, \dots, u_{M-1}) \leq 0$ imposed by the user on the admissible control actions is not too restrictive. An f not satisfying Assumption 1 is for instance $f(u_0, \dots, u_{M-1}) = \|u_0\|^2 + \|u_1\|^2 + \dots + \|u_{M-1}\|^2$, in which case imposing $f(u_0, \dots, u_{M-1}) \leq 0$ leads to the quite contrived situation where $u_0 = 0, u_1 = 0, \dots, u_{M-1} = 0$ is the only admissible control action. More generally, not satisfying Assumption 1 corresponds to a reduction of the degrees of freedom through which we have to accomplish the control task. It is then clear that, unless we are dealing with an over-actuated system, choosing an f not satisfying Assumption 1 would be too restrictive for any sensible control problem. Because of this, Assumption 1 can be regarded as very mild.

Theorem 1 *Fix a value for the confidence parameter β within $(0, 1)$. If the number of extracted disturbance realiza-*

² If $J(\cdot)$ and $L(\cdot)$ are both quadratic and the constraints are quadratic or linear, then the optimization problems (13) and (14) are both quadratic programs and can be solved efficiently.

tions N is chosen so as to satisfy

$$\sum_{i=0}^{d-1} \binom{N}{i} \varepsilon^i (1-\varepsilon)^{N-i} \leq \beta, \quad (16)$$

where d is the dimensionality of (λ, h) , then it holds with confidence at least $1 - \beta$ that

$$\mathbb{P}\{f(\mathbf{u}_{\lambda^*}(\mathbf{w})) \leq 0 \wedge g(\mathbf{x}_{\lambda^*}(\mathbf{w}), \mathbf{u}_{\lambda^*}(\mathbf{w})) \leq h^*\} \geq 1 - \varepsilon,$$

where (λ^*, h^*) is the solution to the cascade of problems (13) and (14).

Theorem 1 states that with high confidence $1 - \beta$ the solution (λ^*, h^*) obtained by solving the scenario cascade of problems (13) and (14) is feasible for the original probabilistic constraint (15) in (10) and (11). Note that the presence of the confidence parameter β is intrinsic and is related to fact that the obtained solution depends on the random extraction $\mathbf{w}^{(1)}, \dots, \mathbf{w}^{(N)}$: β is needed to account for the possibility that a not representative enough sample $\mathbf{w}^{(1)}, \dots, \mathbf{w}^{(N)}$ is seen. However exploiting the results in [1] it can be shown that the number of required samples N according to (16) scales logarithmically with $1/\beta$. Hence β can be chosen to be very small such as 10^{-5} or 10^{-7} without affecting N too much, so that the fact that the achieved solution (λ^*, h^*) satisfies the probabilistic constraint (15) in (10) and (11) can be taken for granted.

Note that N has to be chosen so as to satisfy the inequality (16) for the result in Theorem 1 to hold. In practice, N should be set equal to the lowest value compatible with (16) in order to avoid conservativeness of the solution to the cascade of problems (13) and (14). Alternatively, one could opt for the scenario solution with constraint removal in [12]. The results of Theorem 1 should then be extended to the cascade of scenario problems with constraint removal. This is not straightforward and requires further investigations.

Proof of Theorem 1: For a given (λ, h) , define the violation probability of (λ, h) as

$$\begin{aligned} V(\lambda, h) &:= \mathbb{P}\left\{f(\mathbf{u}_\lambda(\mathbf{w})) > 0 \vee g(\mathbf{x}_\lambda(\mathbf{w}), \mathbf{u}_\lambda(\mathbf{w})) > h\right\} \\ &= 1 - \mathbb{P}\left\{f(\mathbf{u}_\lambda(\mathbf{w})) \leq 0 \wedge g(\mathbf{x}_\lambda(\mathbf{w}), \mathbf{u}_\lambda(\mathbf{w})) \leq h\right\}, \end{aligned}$$

where \vee stands for “or”. Then, Theorem 1 amounts to showing that

$$\mathbb{P}^N\{V(\lambda^*, h^*) > \varepsilon\} \leq \beta, \quad (17)$$

where \mathbb{P}^N is the product probability underlying the independent extraction of the sample $\mathbf{w}^{(1)}, \dots, \mathbf{w}^{(N)}$ based on which the solution (λ^*, h^*) is computed.

Consider the following auxiliary scenario programs

$$\begin{aligned} \min_{\lambda, h \geq 0} L(h) + \frac{1}{n} J(\lambda) \quad \text{subject to:} \quad (18) \\ f(\mathbf{u}_\lambda(\mathbf{w}^{(k)})) \leq 0, k = 1 \dots N, \\ g(\mathbf{x}_\lambda(\mathbf{w}^{(k)}), \mathbf{u}_\lambda(\mathbf{w}^{(k)})) \leq h, k = 1 \dots N, \end{aligned}$$

for $n = 1, 2, \dots$, and denote by (λ_n^*, h_n^*) its optimal solution, which exists and is unique, since: i. the cost function $L(h) + \frac{1}{n} J(\lambda)$ has compact level sets for every $n \geq 1$ thanks to its strict convexity; ii. the optimization feasibility domain defined by the constraints in (18) is close and nonempty. The following two properties hold:

1. For every $n \geq 1$, it holds that

$$\mathbb{P}^N\{V(\lambda_n^*, h_n^*) > \varepsilon\} \leq \beta. \quad (19)$$

2. For every multisample $\mathbf{w}^{(1)}, \dots, \mathbf{w}^{(N)}$, the solution to (18) converges to the solution to (13) and (14), namely,

$$(\lambda_n^*, h_n^*) \rightarrow (\lambda^*, h^*) \text{ as } n \rightarrow \infty. \quad (20)$$

Formal proofs of (19) and (20) are given in Appendix A and Appendix B.

Remark 3 *The auxiliary problem (18) can be interpreted as the sample-based counterpart of the problem (12) considered in [19,21], with $\mu = n$. Property (20) shows that the solution of the cascade of problems (13) and (14) under consideration in this paper can be retrieved in the limit from (18), and we will take advantage of this important theoretical result next to derive the probabilistic guarantee (17). On the other hand, it is important to stress that in no way this allows one to conclude that the cascade of problems (13) and (14) is of no use because one can simply take (18) with n large enough to practically retrieve the solution of (13) and (14). The reason is twofold: i. how large n must be chosen depends significantly on the problem at hand and finding the proper value may require multiple trials and be demanding as compared to the one-shot resolution of the cascade of problems (13) and (14); ii. more importantly, when n is too large, (18) may turn out to be ill-conditioned for a numerical resolution. Indeed, for large values of n , numerical solvers tends to ignore the presence of $J(\lambda)$ and the cost J of the returned solution can then be very different from the cost J of the solution of (13) and (14). These aspects are exemplified next in Section 5 by means of a numerical example.*

We now capitalize on (19) and (20) to show that (17) holds. To this purpose, start by fixing a sample $\mathbf{w}^{(1)}, \dots, \mathbf{w}^{(N)}$ such that $V(\lambda^*, h^*) > \varepsilon$, which, we recall, means that

$$\mathbb{P}\left\{f(\mathbf{u}(\mathbf{w}, \lambda^*)) > 0 \vee g(\mathbf{x}(\mathbf{w}, \lambda^*), \mathbf{u}(\mathbf{w}, \lambda^*)) > h^*\right\} > \varepsilon.$$

By continuity of f and g , this implies that

$$\mathbb{P}\left\{f(\mathbf{u}(\mathbf{w}, \lambda)) > 0 \vee g(\mathbf{x}(\mathbf{w}, \lambda), \mathbf{u}(\mathbf{w}, \lambda)) > h\right\} > \varepsilon,$$

for all $(\lambda, h) : \|(\lambda, h) - (\lambda^*, h^*)\| \leq r$ for a radius r small enough, and, since $(\lambda_n^*, h_n^*) \rightarrow (\lambda^*, h^*)$ so that $\|(\lambda, h) - (\lambda^*, h^*)\| \leq r$ for all n bigger than a suitable \bar{n} , we can conclude that

$$V(\lambda_n^*, h_n^*) > \varepsilon, \quad (21)$$

for $n > \bar{n}$. If we now let $\mathbf{w}^{(1)}, \dots, \mathbf{w}^{(N)}$ vary and we consider the indicator function $\mathbb{I}_{\{\mathbf{w}^{(1)}, \dots, \mathbf{w}^{(N)} : V(\lambda_n^*, h_n^*) > \varepsilon\}}$, then (21) yields

$$\mathbb{I}_{\{V(\lambda^*, h^*) > \varepsilon\}} \cdot \mathbb{I}_{\{V(\lambda_n^*, h_n^*) > \varepsilon\}} \xrightarrow{n \rightarrow \infty} \mathbb{I}_{\{V(\lambda^*, h^*) > \varepsilon\}},$$

for all possible realizations of $\mathbf{w}^{(1)}, \dots, \mathbf{w}^{(N)}$. Note that

$$\begin{aligned} & \lim_{n \rightarrow \infty} \mathbb{P}^N \{V(\lambda_n^*, h_n^*) > \varepsilon\} \\ &= \lim_{n \rightarrow \infty} \int \mathbb{I}_{\{V(\lambda_n^*, h_n^*) > \varepsilon\}} \mathbb{P}^N \{d\mathbf{w}^{(1)}, \dots, d\mathbf{w}^{(N)}\} \\ &\geq \lim_{n \rightarrow \infty} \int \mathbb{I}_{\{V(\lambda^*, h^*) > \varepsilon\}} \cdot \mathbb{I}_{\{V(\lambda_n^*, h_n^*) > \varepsilon\}} \mathbb{P}^N \{d\mathbf{w}^{(1)}, \dots, d\mathbf{w}^{(N)}\}, \end{aligned}$$

where the last step holds because we are restricting the domain of integration. By the Lebesgue dominated convergence theorem ([39, §5.9]), we have

$$\begin{aligned} & \lim_{n \rightarrow \infty} \int \mathbb{I}_{\{V(\lambda^*, h^*) > \varepsilon\}} \cdot \mathbb{I}_{\{V(\lambda_n^*, h_n^*) > \varepsilon\}} \mathbb{P}^N \{d\mathbf{w}^{(1)}, \dots, d\mathbf{w}^{(N)}\} \\ &= \int \mathbb{I}_{\{V(\lambda^*, h^*) > \varepsilon\}} \mathbb{P}^N \{d\mathbf{w}^{(1)}, \dots, d\mathbf{w}^{(N)}\}, \end{aligned}$$

which finally leads to

$$\lim_{n \rightarrow \infty} \mathbb{P}^N \{V(\lambda_n^*, h_n^*) > \varepsilon\} \geq \mathbb{P}^N \{V(\lambda^*, h^*) > \varepsilon\}.$$

Since $\mathbb{P}^N \{V(\lambda_n^*, h_n^*) > \varepsilon\} \leq \beta, \forall n$, relation (17) is proven. ■

5 Numerical Example

In this section we apply the proposed approach to a numerical example inspired by [18].

We consider the mechanical system composed by 4 masses and 4 springs shown in Fig. 1. Masses and stiffness coefficients of springs are all equal to 1. The state $x = [d_1, d_2, d_3, d_4, \dot{d}_1, \dot{d}_2, \dot{d}_3, \dot{d}_4]^T$ is formed by the displacements of masses with respect to nominal positions and their derivatives; the initial state is $x_0 = [0, 3, 0, -3, 0, 0, 0, 0]^T$. The control input $u = [u_1, u_2, u_3]^T$ is instead formed by the forces acting on the masses depicted in Fig. 1. The control action is kept constant over the sampling period $[t, t + T_s)$, with $T_s = 1$

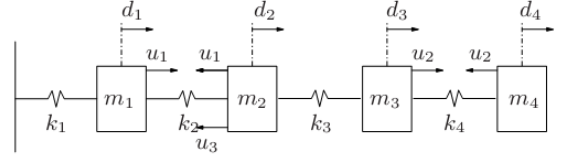


Fig. 1. Scheme of the mechanical system.

s, and we work with the resulting discretized system. A stochastic additive disturbance affecting both the displacement and the speed of the fourth mass is supposed to be also present. This results in a system as in (1), with matrices A and B as obtained from discretization, $B_w = [B_w^1 \ B_w^2]$, $B_w^1 = [0 \ 0 \ 0 \ 1 \ 0 \ 0 \ 0 \ 0]^T$, $B_w^2 = [0 \ 0 \ 0 \ 0 \ 0 \ 0 \ 0 \ 1]^T$, and $w \sim WGN(0, I_2)$.

The control horizon is $M = 5$ and the goal is to keep the spring deformations within a safe range by using the smallest possible control action. To this purpose, we consider the cost function $J = \|\gamma\|^2 + \|\Theta\|_F^2$, where $\|\cdot\|_F$ denotes the Frobenius norm, and the state constraint

$$\left\| \begin{array}{c} x_{1,t} \\ x_{2,t} - x_{1,t} \\ x_{3,t} - x_{2,t} \\ x_{4,t} - x_{3,t} \end{array} \right\|_{\infty} \leq 3, \quad t = 1, \dots, M. \quad (22)$$

To deal with the presence of the disturbance w , the constraint is enforced in probability with $\varepsilon = 0.05$.

Following the approach of Section 3, the optimization variables $h_i, i = 1, \dots, M$, are introduced so as to ensure the feasibility of the probabilistic constraint. We set $\beta = 10^{-6}$ resulting in $N = 2447$ according to (16). Eventually, the cascade of problems (13)-(14) is solved with $L(h) = h^T h$.

In our numerical experiment imposing the original state constraint (22) over the seen scenarios led to infeasibility; indeed, solving problem (13) gave $h_1^* = 0.53, h_2^* = 0.21, h_3^* = 0.38, h_4^* = 0.18$, while $h_5^* = 0$. Notably, the approach succeeds in recognizing that no relaxation is necessary for the constraint at $t = 5$. The cost $J(\lambda^*)$ achieved after solving problem (14) was 56.03. Some Monte-Carlo simulations revealed that the probabilistic constraint (15) was satisfied by the solution (λ^*, h^*) as it is expected given Theorem 1 (the probability of violation turned out to be 0.02). Interestingly enough, though the probabilistic satisfaction of the original constraint (22) could not be guaranteed, it turned out that the probability of violation of the original state constraint (22) was 0.056, which is just slightly above the desired level 0.05. This is so because the proposed approach implements just the minimum extent of constraint relaxation that is needed to recover feasibility.

We now compare the approach of this paper with that of [19,21], by considering a sample-based counterpart of problem (12) (see Remarks 2 and 3). To this aim, Table 1 reports the obtained cost function $J(\lambda^*)$, the values of $h_i^*, i = 1, \dots, M$, as well as the empirical probability of violation

of the original state constraint (22) for the cascade approach of this paper and for the approach of [19,21] for various values of μ . As it appears, the cascade approach of this paper

Table 1

The cascade approach of this paper vs. the approach of [19,21].

Approach	J	h_1	h_2	h_3	h_4	h_5	\hat{V}
cascade	56.03	0.53	0.21	0.38	0.18	0.0	0.055
$\mu = 100$	51.07	0.53	0.22	0.40	0.18	0.08	0.078
$\mu = 10$	16.95	1.01	0.70	0.55	0.28	0.25	0.079
$\mu = 1$	4.88	1.57	1.15	0.89	0.54	0.48	0.133
$\mu = 0.1$	1.14	1.81	2.09	2.47	0.99	1.53	0.442

obtains the smallest relaxations and the best probability of violation of the original constraint. At the same time, the cost J keeps moderate enough thanks to the minimization in (14). As μ increases the results of the approach of [19,21] tend to those of the cascade approach and, in principle, the two approaches are equivalent when $\mu \rightarrow \infty$ (see again Remark 3). However, it is worth noticing that, in line with the observation made in Remark 3, for too large values of μ we were not able to retrieve the solution of the cascade approach, since too little importance was given to $J(\lambda)$ and the numerical solver returned solutions with the same h^* as the cascade approach, but with a very big $J(\lambda^*)$ (in our simulations, this was 526.32). This shows that obtaining a proper tradeoff between $J(\lambda)$ and $L(h)$ by selecting μ is in general difficult, and, anyway, it requires trials and errors.

6 Conclusion

In this paper, we have described a sample-based method for finite horizon constrained control of a linear system affected by a possibly unbounded additive disturbance and we have formally proved that it guarantees probabilistic constraint feasibility. Infeasibility is a relevant issue that can hamper the applicability of MPC strategies. The achievements of this paper then permit one to assemble a recursively feasible MPC scheme where a relaxed optimization problem is solved over a receding horizon. Differently from alternative solutions, relaxation has been suitably designed so as to have a limited impact on control performance.

References

[1] T. Alamo, R. Tempo, A. Luque, and D. Ramirez. Randomized methods for design of uncertain systems: Sample complexity and sequential algorithms. *Automatica*, 52:160 – 172, 2015.

[2] I. Batina, A. Stoorvogel, and S. Weiland. Optimal control of linear, stochastic systems with state and input constraints. In *Proc. of the 41st IEEE Conference on Decision and Control*, Dec. 2002.

[3] D. Bertsimas and D. Brown. Constrained stochastic LQC: A tractable approach. *Automatic Control, IEEE Transactions on*, 52(10):1826–1841, 2007.

[4] L. Blackmore and M. Ono. Convex chance constrained predictive control without sampling. *Proceedings of the AIAA Guidance, Navigation and Control Conference*, 2009.

[5] G. Calafiore and M. Campi. Uncertain convex programs: randomized solutions and confidence levels. *Mathematical Programming*, 102(1):25–46, 2005.

[6] G. Calafiore and M. Campi. The scenario approach to robust control design. *IEEE Transactions on Automatic Control*, 51(5):742–753, 2006.

[7] G. Calafiore and L. Fagiano. Robust model predictive control via random convex programming. In *Decision and Control and European Control Conference (CDC-ECC), 2011 50th IEEE Conference on*, pages 1910–1915, 2011.

[8] G. Calafiore and L. Fagiano. Robust model predictive control via scenario optimization. *IEEE Transactions on Automatic Control*, 58(1):219–224, 2013.

[9] E. Camacho and C. Bordons. *Model Predictive Control*. Springer, London, 2004.

[10] M. Campi and A. Carè. Random convex programs with l_1 -regularization: sparsity and generalization. *SIAM Journal on Control and Optimization*, 51(5):3532–3557, 2013.

[11] M. Campi and S. Garatti. The exact feasibility of randomized solutions of uncertain convex programs. *SIAM Journal on Optimization*, 19(3):1211–1230, 2008.

[12] M. Campi and S. Garatti. A sampling-and-discarding approach to chance-constrained optimization: Feasibility and optimality. *Journal of Optimization Theory and Applications*, 148(2):257–280, 2011.

[13] M. Campi and S. Garatti. Wait-and-judge scenario optimization. *Mathematical Programming*, 167(1):155–189, 2018.

[14] M. Campi, S. Garatti, and M. Prandini. The scenario approach for systems and control design. *Annual Reviews in Control*, 33(2):149–157, 2009.

[15] M. Cannon, B. Kouvaritakis, S. Rakovic, and Q. Cheng. Stochastic tubes in model predictive control with probabilistic constraints. *IEEE Transactions on Automatic Control*, 56(1):194–200, 2011.

[16] A. Carè, S. Garatti, and M. Campi. Scenario min-max optimization and the risk of empirical costs. *SIAM Journal on Optimization*, 25(4):2061–2080, 2015.

[17] Q. Cheng, M. Cannon, B. Kouvaritakis, and M. Evans. Stochastic MPC for systems with both multiplicative and additive disturbances. In *Proceedings of the 19th IFAC World Congress*, Cape Town, South Africa, 2014.

[18] E. Cinquemani, M. Agarwal, D. Chatterjee, and J. Lygeros. Convexity and convex approximations of discrete-time stochastic control problems with constraints. *Automatica*, 47(9):2082–2087, 2011.

[19] L. Deori, S. Garatti, and M. Prandini. Stochastic constrained control: Trading performance for state constraint feasibility. In *Proceedings of the 2013 European Control Conference*, pages 2740–2745, Zurich, Switzerland, 2013.

[20] L. Deori, S. Garatti, and M. Prandini. Stochastic control with input and state constraints: a relaxation technique to ensure feasibility. In *Proceedings of the 54th IEEE Conference on Decision and Control*, Osaka, Japan, 2015.

[21] L. Deori, S. Garatti, and M. Prandini. Trading performance for state constraint feasibility in stochastic constrained control: a randomized approach. *Journal of the Franklin Institute*, 354(354):501–529, 2017.

[22] M. Farina, L. Giulioni, L. Magni, and R. Scattolini. An mpc approach to output-feedback control of stochastic linear discrete-time systems. *IEEE Transactions on Automatic Control*, 55:140–149, 2015.

[23] M. Farina, L. Giulioni, and R. Scattolini. Stochastic linear model predictive control with chance constraints - a review. *Journal of Process Control*, 44:53–67, 2016.

[24] P. Goulart, E. Kerrigan, and J. Maciejowski. Optimization over state feedback policies for robust control with constraints. *Automatica*, 42(4):523–533, April 2006.

- [25] D. V. Hessem and O. Bosgra. Stochastic closed-loop model predictive control of continuous nonlinear chemical processes. *Journal of Process Control*, 16(3):225 – 241, 2006.
- [26] P. Hokayem, E. Cinquemani, D. Chatterjee, F. Ramponi, and J. Lygeros. Stochastic receding horizon control with output feedback and bounded controls. *Automatica*, 48(1):77 – 88, 2012.
- [27] B. Kouvaritakis and M. Cannon. *Model Predictive Control*. Springer, 2016.
- [28] M. Lorenzen, F. Dabbene, R. Tempo, and F. Allgöwer. Constraint-tightening and stability in stochastic model predictive control. *IEEE Transactions on Automatic Control*, 62(7):3165–3177, July 2017.
- [29] J. Maciejowski. *Predictive Control with constraints*. Prentice Hall, Harlow, England, 2002.
- [30] K. Margellos, M. Prandini, and J. Lygeros. On the connection between compression learning and scenario based single-stage and cascading optimization problems. *IEEE Transactions on Automatic Control*, 60(10):2716–2721, Oct 2015.
- [31] M. Ono and B. Williams. Iterative risk allocation: A new approach to robust model predictive control with a joint chance constraint. In *Proc. of the 47th IEEE Conference on Decision and Control*, Dec. 2008.
- [32] M. Prandini, Garatti, S., and J. Lygeros. A Randomized Approach to Stochastic Model Predictive Control. In *IEEE Conference on Decision and Control*, Maui, Hawaii, USA, Dec. 2012.
- [33] J. Primbs. A soft constraint approach to stochastic receding horizon control. In *Proc. of the 46th IEEE Conference on Decision and Control*, Dec. 2007.
- [34] J. Primbs and H. Chang. Stochastic receding horizon control of constrained linear systems with state and control multiplicative noise. *Automatic Control, IEEE Transactions on*, 54(2):221–230, 2009.
- [35] J. Rawlings and D. Mayne. *Model Predictive Control Theory and Design*. Nob Hill Publishing, Madison, WI, 2009.
- [36] G. Schildbach, G. Calafiore, L. Fagiano, and M. Morari. Randomized Model Predictive Control for Stochastic Linear Systems. In *American Control Conference*, pages 417–422, Montreal, Canada, June 2012.
- [37] G. Schildbach, L. Fagiano, C. Frei, and M. Morari. The Scenario Approach for Stochastic Model Predictive Control with Bounds on Closed-Loop Constraint Violations. *Automatica*, 50(12):3009–3018, 2014.
- [38] G. Schildbach, L. Fagiano, and M. Morari. Randomized solutions to convex programs with multiple chance constraints. *SIAM Journal on Optimization*, 23(4):2479–2501, 2013.
- [39] D. Williams. *Probability with Martingales*. Cambridge mathematical textbooks. Cambridge University Press, 1991.
- [40] X. Zhang, S. Grammatico, K. Margellos, P. Goulart, and J. Lygeros. Randomized nonlinear MPC for uncertain control-affine systems with bounded closed-loop constraint violations. In *Proceedings of the 19th IFAC World Congress*, Cape Town, South Africa, 2014.
- [41] X. Zhang, S. Grammatico, G. Schildbach, P. Goulart, and J. Lygeros. On the sample size of random convex programs with structured dependence on the uncertainty. *Automatica*, 60:182–188, 2015.

A Proof of (19)

By adding a slack variable $v \in \mathbb{R}$, problem (18) can be rewritten in epigraphic form as:

$$\begin{aligned} \min_{\lambda, h \geq 0, v} v \text{ subject to:} & \quad (\text{A.1}) \\ f(\mathbf{u}_\lambda(\mathbf{w}^{(k)})) \leq 0, k = 1 \dots N, & \\ g(\mathbf{x}_\lambda(\mathbf{w}^{(k)}), \mathbf{u}_\lambda(\mathbf{w}^{(k)})) \leq h, k = 1 \dots N, & \\ L(h) + \frac{1}{n}J(\lambda) \leq v. & \end{aligned}$$

The solution to problem (A.1) is still unique, and the assumptions of Theorem 2.4 in [11] are satisfied thanks to Assumption 1. An application of this theorem gives

$$\mathbb{P}^N \{V(\lambda_n^*, h_n^*) > \varepsilon\} \leq \sum_{i=0}^d \binom{N}{i} \varepsilon^i (1 - \varepsilon)^{N-i},$$

where we have d in place of $d - 1$ because in (A.1) the number of optimization variables has been augmented by 1 and is equal to $d + 1$. On the other hand, since the slack variable v does not enter the expression defining it, the constraint

$$\{\lambda, h, v : f(\mathbf{u}_\lambda(\mathbf{w})) \leq 0 \wedge g(\mathbf{x}_\lambda(\mathbf{w}), \mathbf{u}_\lambda(\mathbf{w})) \leq h\}$$

is, irrespective of \mathbf{w} , a cylindroid infinitely extended along the v direction. This entails that the family (with respect to the variability of \mathbf{w}) of constraints above has a so-called support rank equal to d , according to Definition 3.6 of [38] (see also [41]). The conclusion that

$$\mathbb{P}^N \{V(\lambda_n^*, h_n^*) > \varepsilon\} \leq \sum_{i=0}^{d-1} \binom{N}{i} \varepsilon^i (1 - \varepsilon)^{N-i}$$

then follows by invoking the observation made in [38] that Theorem 2.4 of [11] still applies by replacing the optimization domain dimensionality with the support rank (see Lemma 3.8). ■

B Proof of (20)

To show that $(\lambda_n^*, h_n^*) \rightarrow (\lambda^*, h^*)$ as $n \rightarrow \infty$, consider the sets

$$\begin{aligned} \mathcal{H}_n = \left\{ (\lambda, h) : (\lambda, h) \text{ is feasible for (18) and} \right. \\ \left. L(h) + \frac{1}{n}J(\lambda) \leq L(h^*) + \frac{1}{n}J(\lambda^*) \right\}, \end{aligned}$$

for $n = 1, 2, \dots$. In words, n by n , \mathcal{H}_n is the set of all feasible points for (18) that also belong to the smallest level set of the cost function of (18) containing the solution (λ^*, h^*) of (13) and (14). Note that, while the level set changes with n , the feasibility domain of (18) remains the same for all n and it coincides with the feasibility domain of (13). This entails

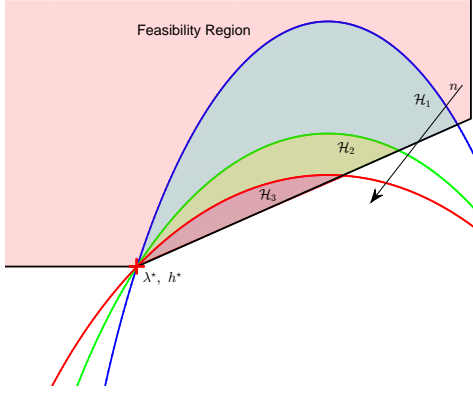


Fig. B.1. The sets \mathcal{H}_n 's in a simple case ($h, \lambda \in \mathbb{R}$, $L(h) = h^2$, $J(\lambda) = 3\lambda^2$).

that (λ^*, h^*) belongs to \mathcal{H}_n for all n , showing also that \mathcal{H}_n is nonempty. Moreover, n by n , we have that

$$(\lambda_n^*, h_n^*) \in \mathcal{H}_n, \quad (\text{B.1})$$

because (λ_n^*, h_n^*) is feasible for (18), and, being also optimal, its cost value must be better than that of (λ^*, h^*) , which is the second condition defining \mathcal{H}_n .

A fundamental property of the family of sets \mathcal{H}_n is that

$$\mathcal{H}_1 \supseteq \mathcal{H}_2 \supseteq \dots \supseteq \mathcal{H}_n \supseteq \mathcal{H}_{n+1} \supseteq \dots, \quad (\text{B.2})$$

as pictorially depicted in Fig. B.1. To show (B.2), suppose that a $(\bar{\lambda}, \bar{h})$ belongs to \mathcal{H}_{n+1} . From

$$L(\bar{h}) + \frac{1}{n+1}J(\bar{\lambda}) \leq L(h^*) + \frac{1}{n+1}J(\lambda^*)$$

it follows that $J(\bar{\lambda}) \leq (n+1)(L(h^*) - L(\bar{h})) + J(\lambda^*)$. Whence,

$$\begin{aligned} L(\bar{h}) + \frac{1}{n}J(\bar{\lambda}) &\leq L(\bar{h}) + \frac{n+1}{n}(L(h^*) - L(\bar{h})) + \frac{1}{n}J(\lambda^*) \\ &= L(h^*) + \frac{1}{n}(L(h^*) - L(\bar{h})) + \frac{1}{n}J(\lambda^*) \\ &\leq L(h^*) + \frac{1}{n}J(\lambda^*), \end{aligned}$$

where the last inequality follows because $L(h^*) - L(\bar{h}) \leq 0$ being $L(h^*)$ the lowest among feasible points by the definition of h^* . This shows that $(\bar{\lambda}, \bar{h}) \in \mathcal{H}_n$ too, that is, (B.2) holds.

From (B.1) and (B.2), we have that $(\lambda_n^*, h_n^*) \in \mathcal{H}_1, \forall n$. Set \mathcal{H}_1 is compact, being the intersection of the feasibility domain of (13), which is close, with a level set of $L(h) + \frac{1}{n}J(\lambda)$, which is compact thanks to the assumptions of strict of convexity of L and J . It then follows that the sequence (λ_n^*, h_n^*) have limit points, which are feasible for (13). For simplicity, assume that there is just one, say $(\lambda_\infty^*, h_\infty^*)$, so that the sequence (λ_n^*, h_n^*) is convergent to $(\lambda_\infty^*, h_\infty^*)$. If not, simply

repeat the argument that follows to each limit point and the corresponding convergent subsequence.

From (B.1) and the definition of \mathcal{H}_n , we have that

$$L(h_n^*) \leq L(h^*) + \frac{1}{n}[J(\lambda^*) - J(\lambda_n^*)],$$

which in turn implies that

$$L(h_\infty^*) = \lim_{n \rightarrow \infty} L(h_n^*) \leq L(h^*) + \lim_{n \rightarrow \infty} \frac{1}{n}[J(\lambda^*) - J(\lambda_n^*)] = L(h^*).$$

Yet, being $L(h^*)$ minimal, it cannot be that a strict inequality holds, so that eventually $L(h_\infty^*) = L(h^*)$. If $h_\infty^* \neq h^*$, then $(\frac{1}{2}\lambda^* + \frac{1}{2}\lambda_\infty^*, \frac{1}{2}h^* + \frac{1}{2}h_\infty^*)$ would be feasible for (13) thanks to the convexity of the feasible domain, while the strict convexity of $L(h)$ would give

$$L\left(\frac{1}{2}h^* + \frac{1}{2}h_\infty^*\right) < \frac{1}{2}L(h^*) + \frac{1}{2}L(h_\infty^*) = L(h^*),$$

so contradicting the minimality of $L(h^*)$. Hence, $h_\infty^* = h^*$.

From $(\lambda_n^*, h_n^*) \in \mathcal{H}_1$, we have that $J(\lambda_n^*) \leq L(h^*) - L(h_n^*) + J(\lambda^*)$ which, taking the limit, gives

$$J(\lambda_\infty^*) \leq \lim_{n \rightarrow \infty} L(h^*) - L(h_n^*) + J(\lambda^*) = J(\lambda^*).$$

Plainly, it must be that $J(\lambda_\infty^*) = J(\lambda^*)$, for, otherwise, being λ_∞^* feasible for (14), $J(\lambda_\infty^*) < J(\lambda^*)$ would contradict the minimality of $J(\lambda^*)$. Moreover, if $\lambda_\infty^* \neq \lambda^*$, then $\frac{1}{2}\lambda^* + \frac{1}{2}\lambda_\infty^*$ would be feasible for (14), and, because of the strict convexity of $J(\lambda)$ we would have

$$J\left(\frac{1}{2}\lambda^* + \frac{1}{2}\lambda_\infty^*\right) < \frac{1}{2}J(\lambda^*) + \frac{1}{2}J(\lambda_\infty^*) = J(\lambda^*),$$

contradicting again the minimality of $J(\lambda^*)$. Hence, $\lambda_\infty^* = \lambda^*$, and this concludes the proof of Property 2. ■