Contents

7 Stochastic Linear Control

In this section we consider the stochastic reference tracking problem. It is required to move the states and manipulated variables of the system, shown in (85), to the set point (x_{sp}, u_{sp}) by manipulating the input variables u.

$$x_{t+1} = f(x_t, u_t) + w_{t+1}$$
$$y_{t+1} = g(x_{t+1}) + v_{t+1}$$

(85)

We assume uncorrelated zero mean additive Gaussian noise in both the state function f and the observation function g with known covariances W and V respectively. Clearly it is not possible to achieve perfect control (zero offset at steady state) because of the noise terms, specifically w_t. For this reason we need to relax the set point goal a little bit. We will be content if our controller is able to achieve Definition 7.1.

Definition 7.1. Stochastic Reference Tracking Goal: Suppose we have designed a controller and set $\delta > 0$ as a controller benchmark. If there exists some positive number $t^* < \infty$ such that $\forall~t > t^*$ the controller input causes $\mathbb{E}[(x_t - x_{sp})^T Q(x_t - x_{sp}) + (u_t - u_{sp})^T R(u_t - u_{sp})] < \delta$ we will have satisfied the Stochastic Reference Tracking Goal given δ.

In this section we limit ourselves by only considering controllers developed using a single linear model of the underlying, possibly nonlinear, system functions f and g. The linearised model control is based upon is shown in (86) and is subject to the same noise as (85).

$$x_{t+1} = Ax_t + Bu_t + w_{t+1}$$
$$y_{t+1} = Cx_{t+1} + v_{t+1}$$

(86)

We will endeavour to develop predictive controllers using the Graphical Models of Section 5 and 6.

7.1 Current Literature

Linear unconstrained stochastic control subject to white additive Gaussian noise is well studied in literature. The Linear Quadratic Gaussian (LQG) controller introduced in Section 2.4 is one of the most fundamental results in Optimal Control Theory [10]. A significant drawback of the LQG controller, and by extension the LQR controller, is that it is inherently unconstrained.

Conventional deterministic MPC is very well studied in literature [37] and can be seen as the constrained generalisation of the LQR controller. A further generalisation of deterministic MPC is stochastic MPC whereby either the variables, the constraints or both have stochastic elements. In current literature the trend is to convert all the stochastic elements of the control problem into deterministic ones. This usually makes the problem somewhat more tractable from a computational point of view.

This conversion is usually achieved via two distinct approaches. In the first approach, which is also the one we employ, the probability distributions are assumed Gaussian and the systems linear. This allows one to greatly simplify the problem at the cost of those relatively strong assumptions. The second approach is to use a particle/sampling approach. Here the distributions are approximated by particles (much like the Particle Filter of Section 6) and no assumptions are made of form of the distributions. It is also not necessary to assume linear dynamics. The major practical drawback of this approach is that it can quickly become computationally intractable.

Indeed, this is the approach taken by [6]. They attempt to solve the stochastic MPC problem with stochastic (chance) constraints and variables by approximating the current and predicted distributions with particles. Referring back to Section 6.3 we see that this is indeed possible. By using the particle approach to model distributions it is possible to convert the resultant stochastic optimisation problem into a deterministic one. In the case where linear dynamics are used this becomes a Mixed Integer Linear or Quadratic Programming problem depending on the objective function. Their algorithm is appealing because it is not necessary to assume Gaussian distributions. However, it is possible that the algorithm could become computationally intractable due to the integer constraints which are used to approximate the chance constraints: it is necessary to include an integer variable for each particle at each time step in the prediction horizon. For large problems with long prediction horizons this can be problematic.

The approach taken by [28] is related to the sampling approach. They convert the stochastic chance constrained optimisation problem into a deterministic nonlinear optimisation problem. They then use a simulation approach to ensure that the chance constraints are satisfied. The approach taken by [4] uses a randomized optimisation algorithm in concert with the empirical mean of the variables. A penalty method is used to enforce constraints. The latter paper takes a step in the direction of simplifying the problem without sampling.

In the paper [27] the stochastic variables are assumed to be Gaussian and the stochastic optimisation problem is transformed into a nonlinear optimisation problem. Using the Gaussian assumption they are able to ensure feasibility and constraint satisfaction albeit conservatively. In [31] the feasibility of stochastically constrained predictive control is considered. An algorithm enforcing joint chance constraints and recursive feasibility is discussed using a risk allocation approach.

While [38] mainly concerns stochastic parameters in the optimisation problem it is shown that chance constraints can, in theory, be rewritten as deterministic constraints if the probability distributions are known and affine constraints are used. In [41] and [40] an ellipsoidal approximation is used to ensure constraint satisfaction. Using the ellipsoidal approximations the stochastic optimisation problem can transformed into a second order conic optimisation problem. The approach of using confidence ellipsoids is refined in [7].

Although [43] and [44] primarily deal with a univariate problem they show that if the under-

lying system is linear and Gaussian it is possible to manipulate the constrained stochastic problem into a deterministic problem without any approximations. Their analysis allows the stochastic objective function and constraints to be rewritten as deterministic expressions using the properties of conditionally linear Gaussian distributions.

In this work we show that by starting the MPC design from the graphical model shown in Figure 28 it is possible to easily and intuitively re-derive the LQG controller. During that derivation we naturally arrive at the conclusions reached by [43] and [44]. Next we generalise our method to incorporate stochastic constraints. We use a method related to the ellipsoidal approximation technique to handle the stochastic constraints. We show that our method replaces the chance constraints with linear constraints. This is highly desirable because it allows one to rewrite the full stochastic MPC problem as a conventional deterministic Quadratic Programming MPC problem. All of these results are underpinned by the realisation that Graphical Models, as used in the field of Machine Learning, are closely related to predictive control.

7.2 Unconstrained Stochastic Control

Our goal is to solve the problem in (87) given the current state estimate x_0. If the system is controllable and linear the Stochastic Reference Tracking Goal will be satisfied.

$$
\begin{aligned}
\min_{\mathbf{u}} \ & J_{LQG}(x_0, \mathbf{u}) = \mathbb{E}\left[\frac{1}{2}\sum_{k=0}^{N-1}\left(x_k^T Q x_k + u_k^T R u_k\right) + \frac{1}{2}x_N^T P_f x_N\right] \\
& \text{subject to } x_{t+1} = A x_t + B u_t + w_t \\
& \text{and } y_t = C x_t + v_t
\end{aligned}
\tag{87}
$$

Note that the future inputs $\mathbf{u} = (u_0, u_1, ..., u_{T-1})$ are denoted in boldface to emphasise that it could be a vector of vectors. Inspecting (87) we see that this is none other than the LQG control problem of Section 2.4.2. Therefore we know what the optimal solution should look like.

We start out analysis using the results of Section 5. We immediately realise that the optimal linear state estimator is the Kalman Filter. We assume that at every sequential time step we have the current state estimate, supplied by the Kalman Filter, and denote this by x_0. Since we are using the Kalman Filter the mean and covariance of the state estimate is well defined.

Inspecting **insert GM pic** we note that the state prediction equations derived in Subsection 5.2 are applicable. Thus we can predict the state distributions given the future inputs $\mathbf{u}$.

Before we proceed we prove a very intuitive result in Theorem 7.1. We will use this to link the predictive controller derived using the results of Section 5 to the LQR controller derived in Section 2.4.1. We provide two proofs, the second of which is more general than the first.

Theorem 7.1. Unconstrained Optimisation Equivalence Suppose we have two real valued objective functions $f(x_0, \mathbf{u})$ and $g(x_0, \mathbf{u})$ and we are required to minimise them

with respect to $\mathbf{u}$ over the same space where they are both defined: $\mathbf{u} \in \mathcal{U}$ and $x_0 \in \mathcal{X}$. Furthermore, suppose there exists a real number $k \geq 0$ such that $\forall \mathbf{u} \in \mathcal{U}$ we have that $g(x_0, \mathbf{u}) + k = f(x_0, \mathbf{u})$. Finally, assume the existence and uniqueness of the global minimiser for each problem. Then the global minimiser $\mathbf{u}^*$ of $g(x_0, \mathbf{u})$ also minimises $f(x_0, \mathbf{u})$.

Proof. This proof only holds over functions which are at least twice differentiable and convex. By assumption we know that $\mathbf{u}^*$ is the minimiser of $g(x_0, \mathbf{u})$ given x_0. By the necessary conditions for optimality [18] we know that $\nabla g(x_0, \mathbf{u}^*) = 0$ and that $\nabla^2 g(x_0, \mathbf{u}^*)$ is positive semi-definite. Since f and g are both twice differentiable and $g(x_0, \mathbf{u}^*) + k = f(x_0, \mathbf{u}^*)$ it must hold that $\nabla g(x_0, \mathbf{u}^*) = \nabla f(x_0, \mathbf{u}^*)$ and $\nabla^2 g(x_0, \mathbf{u}^*) = \nabla^2 f(x_0, \mathbf{u}^*)$. Since $\nabla^2 g(x_0, \mathbf{u}^*)$ is positive semi-definite it must be that $\nabla^2 f(x_0, \mathbf{u}^*)$ is also positive semi-definite. Therefore $\mathbf{u}^*$ is necessarily a minimum of f. Since f is convex the minimum must also be a global minimum. $\qquad\square$

Proof. This proof hold over differentiable and non-differentiable objective functions which are not necessarily convex. Suppose not i.e. there exists $\mathbf{u}_g \in \mathcal{U}$ such that $g(x_0, \mathbf{u}_g) < g(x_0, \mathbf{u})$ $\forall \mathbf{u} \in \mathcal{U}$ but $f(x_0, \mathbf{u}_g) \not< f(x_0, \mathbf{u})$ $\forall \mathbf{u} \in \mathcal{U}$. This implies that for $\mathbf{u}_f \in \mathcal{U}$ the global minimiser of f we have $f(x_0, \mathbf{u}_f) \leq f(x_0, \mathbf{u}_g)$.

Consider the case where $f(x_0, \mathbf{u}_f) = f(x_0, \mathbf{u}_g)$. This implies that both $\mathbf{u}_f$ and $\mathbf{u}_g$ are global minimisers of f and contradicts our assumption that the global minimiser is unique.

Consider the case where $f(x_0, \mathbf{u}_f) < f(x_0, \mathbf{u}_g)$. Since $g(x_0, \mathbf{u}) + k = f(x_0, \mathbf{u})$ $\forall \mathbf{u} \in \mathcal{U}$ this implies that $g(x_0, \mathbf{u}_f) < g(x_0, \mathbf{u}_g)$. But this contradicts our assumption that $\mathbf{u}_g$ is the global minimiser of g.

It must then hold that $f(x_0, \mathbf{u}_g) < f(x_0, \mathbf{u})$ $\forall \mathbf{u} \in \mathcal{U}$. Therefore the global minimiser $\mathbf{u}_g$ of $g(x_0, \mathbf{u})$ also minimises $f(x_0, \mathbf{u})$. $\qquad\square$

Now we are in a position to show the equivalence between the LQR control problem and the LQG control problem using the results of Section 5. Theorem 7.2 shows how this is possible. It is quite reassuring to note that by starting within the framework of Graphical Models we arrive at the most important contribution of [43] and [44] in an intuitively simple manner.

Theorem 7.2. LQR and LQG Objective Function Difference Consider the LQR and LQG Objective Functions in (88) and (89) respectively.

$$J_{LQR}(x_0, \mathbf{u}) = \frac{1}{2} \sum_{k=0}^{N-1} \left(x_k^T Q x_k + u_k^T R u_k \right) + \frac{1}{2} x_N^T P_f x_N \tag{88}$$

with $x_{t+1} = A x_t + B u_t$

$$J_{LQG}(x_0, \mathbf{u}) = \mathbb{E} \left[\frac{1}{2} \sum_{k=0}^{N-1} \left(x_k^T Q x_k + u_k^T R u_k \right) + \frac{1}{2} x_N^T P_f x_N \right] \tag{89}$$

with $x_{t+1} = A x_t + B u_t + w_{t+1}$

Suppose x_0 is the state estimate supplied by the Kalman Filter given the latest observation in the stochastic case. In the deterministic case we have that $x_0 = \mathbb{E}[x_0] = \mu_0$ because we exactly observe the state. Given any input sequence $\mathbf{u} \in \mathcal{U}$, where $\mathcal{U}$ is the shared admissible input space, we have that $J_{LQR}(x_0, \mathbf{u}) + \frac{1}{2}\sum_{k=0}^{N} \text{tr}(Q\Sigma_k) = J_{LQG}(x_0, \mathbf{u})$ where $\Sigma_{t+1} = W + A\Sigma_t A^T$ and Σ_0 is the covariance matrix of the current state given by the Kalman Filter.

Proof. Expanding the LQG objective function and noting that $\mathbf{u}$ is deterministic we have (90). Note that the conditional expectations in the expansion originate from the graphical model in Figure 23 (due to the first order Markov assumption).

$$
\begin{aligned}
J_{LQG}(x_0, \mathbf{u}) &= \frac{1}{2}\mathbb{E}\left[x_0^T Q x_0 + u_0^T R u_0\right] + \frac{1}{2}\mathbb{E}\left[x_1^T Q x_1 + u_1^T R u_1 | x_0\right] + ... \\
&+ \frac{1}{2}\mathbb{E}\left[x_{N-1}^T Q x_{N-1} + u_{N-1}^T R u_{N-1} | x_{N-2}\right] + \frac{1}{2}\mathbb{E}\left[x_N^T P_f x_N | x_{N-1}\right] \\
&= \frac{1}{2}\mathbb{E}\left[x_0^T Q x_0\right] + \frac{1}{2}u_0^T R u_0 + \frac{1}{2}\mathbb{E}\left[x_1^T Q x_1 | x_0\right] + \frac{1}{2}u_1^T R u_1 + ... \\
&+ \frac{1}{2}\mathbb{E}\left[x_{N-1}^T Q x_{N-1} | x_{N-2}\right] + \frac{1}{2}u_{N-1}^T R u_{N-1} + \frac{1}{2}\mathbb{E}\left[x_N^T P_f x_N | x_{N-1}\right]
\end{aligned}
\tag{90}
$$

We know that $x_0 \sim \mathcal{N}(\mu_0, \Sigma_0)$ because the current state estimate comes from the Kalman Filter. This means that we can evaluate the first expected value in (90) using Theorem 2.5 as shown in (91).

$$
\mathbb{E}\left[x_0^T Q x_0\right] = \text{tr}(Q\Sigma_0) + \mu_0^T Q \mu_0
\tag{91}
$$

Now we turn out attention to the second expected value in (90). First note that because we have x_0 and $\mathbf{u}$ we can use the result from Section 5.2 to predict (optimally) the distribution of x_1. Therefore we know that $x_1 \sim \mathcal{N}(A\mu_0 + Bu_0, W + A\Sigma_0 A^T)$. Now we let $\mu_1 = A\mu_0 + Bu_0$ and $\Sigma_0 = W + A\Sigma_0 A^T$. Then by using Theorem 2.5 as before we have (92).

$$
\mathbb{E}\left[x_1^T Q x_1 | x_0\right] = \text{tr}(Q\Sigma_1) + \mu_1^T Q \mu_1
\tag{92}
$$

Note that $\text{tr}(Q\Sigma_1)$ does not depend on u_0 but only on the initial state estimate x_0 which is independent of the future inputs $\mathbf{u}$. Notice that we can continue in this manner to simplify the LQG objective function to (93).

$$
\begin{aligned}
J_{LQG}(x_0, \mathbf{u}) &= \frac{1}{2}\sum_{k=0}^{N-1}\left(\mu_k^T Q \mu_k + u_k^T R u_k\right) + \frac{1}{2}\mu_N^T P_f \mu_N + \frac{1}{2}\sum_{k=0}^{N}\text{tr}(Q\Sigma_k) \\
&\text{with } \mu_{t+1} = A\mu_t + Bu_t \\
&\text{and } \Sigma_{t+1} = W + A\Sigma_t A^T
\end{aligned}
\tag{93}
$$

Now note that except for the last term $J_{LQG}(x_0, \mathbf{u})$ is exactly the same as $J_{LQR}(x_0, \mathbf{u})$. The conclusion follows because $\frac{1}{2}\sum_{k=0}^{N}\text{tr}(Q\Sigma_k)$ is independent of $\mathbf{u}$. $\qquad\square$

Finally we combine Theorem 7.1 and 7.2 to produce Theorem 7.3 which is the main result of this subsection.

Theorem 7.3. Solution of the Finite Horizon LQG control problem We wish to solve the LQG control problem within the framework of Graphical Models. The full problem is shown in (94).

$$\min_{\mathbf{u}} V(x_0, \mathbf{u}) = \mathbb{E}\left[\frac{1}{2}\sum_{k=0}^{N-1}\left(x_k^T Q x_k + u_k^T R u_k\right) + \frac{1}{2}x_N^T P_f x_N\right]$$

subject to $x_{t+1} = A x_t + B u_t + w_t$

and $y_t = C x_t + v_t$
$$\tag{94}$$

We assume that we have the Kalman Filter state estimate for x_0. We use Theorem 7.2 to prove that given x_0 and $\forall \mathbf{u} \in \mathcal{U}$ we have that $J_{LQR}(x_0, \mathbf{u}) + \frac{1}{2}\sum_{k=0}^{N}\text{tr}(Q\Sigma_k) = J_{LQG}(x_0, \mathbf{u})$ with $\frac{1}{2}\sum_{k=0}^{N}\text{tr}(Q\Sigma_k) \in \mathbb{R}$ a constant depending only on x_0. Thus we can use Theorem 7.1 to prove that we only need to solve for the optimal controller input $\mathbf{u}^0$ using the LQR objective function. Thus we can use Theorem 2.10 to find $\mathbf{u}$.

As we have mentioned before, the Separation Theorem implies that the solution of the LQG control problem is achieved by using the Kalman Filter to optimally estimate the current state and then using that state estimate in the optimal LQR controller. It is reassuring that Theorem 7.3 is confirmed by this result. The primary benefit of the Graphical Model approach is clear: we have solved the LQG problem without resorting to Stochastic Dynamical Programming.

Under some circumstances it is also possible to extend the result of Theorem 7.3 to the infinite horizon case as shown in Theorem 7.4.

Theorem 7.4. Solution of the Infinite Horizon LQG control problem If the linear model of (86) is stable then, using, with some minor adjustments, Theorems 7.1 and 7.2 it is possible to show that the infinite horizon LQG problem is solved in a similar manner: the Kalman Filter state estimate is used in conjunction with the infinite horizon LQR solution. This result can also be obtained by using the Separation Theorem.

To clarify why it is important that the linear system, i.e. the matrix A, is stable, consider the quantity $\frac{1}{2}\sum_{k=0}^{N}\text{tr}(Q\Sigma_k)$. If it is unbounded the optimisation minimum will tend to infinity. Inspecting $\Sigma_{t+1} = W + A\Sigma_t A^T$ we see that $||\Sigma_\infty||$ will be unbounded if $||A\Sigma_t A^T||$ becomes unbounded (W is a constant). Note that $||\cdot||$ is some matrix norm. It can be shown that if the eigenvalues of A are less than unity i.e. the linear model is stable, then $||A\Sigma_{t+1}A^T|| \leq ||A\Sigma_{t+1}A^T||$ which implies that $\frac{1}{2}\sum_{k=0}^{N}\text{tr}(Q\Sigma_k)$ is bounded and the optimisation is reasonable.

7.3 Constrained Stochastic Control

The goal of this section is to solve the stochastic constrained optimisation problem shown in (95). We assume that the underlying system is linear and the probability distributions are

Gaussian[5]. We also restrict our analysis to affine constraints.

$$\min_{\mathbf{u}} \mathbb{E}\left[\frac{1}{2}\sum_{k=0}^{N-1}\left(x_k^T Q x_k + u_k^T R u_k\right) + \frac{1}{2}x_N^T P_f x_N\right]$$

subject to $x_{t+1} = Ax_t + Bu_t + w_t$

and $y_t = Cx_t + v_t$

and $\mathbb{E}[dx_t + e] \geq 0 \ \forall \ t = 1, ..., N$

and $\Pr(dx_t + e \geq 0) \geq p \ \forall \ t = 1, ..., N$

(95)

It might seem that the last constraint is a duplicate of the preceding one. Closer inspection reveals their different character. The first inequality constraint requires that the predicted states satisfy the constraint "on average" while the second inequality constraint requires that the predicted states jointly satisfy the constraint with at least some probability p.

Theorem 7.5. Affine Expected Value Constraints Suppose we have a stochastic variable x with a known Gaussian distribution. Then the stochastic constraint $\mathbb{E}[dx + e] \geq 0$ simplifies to the deterministic constraint $d\mu + e \geq 0$ where $\mathbb{E}[x] = \mu$ is the mean of the stochastic variable.

Proof. We know that x is a Gaussian stochastic variable. By the results of Section 2.5 we know that $\mathbb{E}[dx + e] = d\mu + e$. This immediately implies the Theorem. $\square$

Theorem 7.6. Shortest Squared Mahalanobis Distance between a Hyperplane and a Point Suppose we are given a symmetric positive semi-definite matrix S and a point y. The shortest squared Mahalanobis Distance between y and the hyperplane $b^T x + c = 0$ is given by $\frac{(b^T y + c)^2}{b^T S b}$.

Proof. It is natural to formulate Theorem 7.6 as an optimisation problem as shown in (96).

$$\min_{x} (x - y)^T S^{-1}(x - y)$$

subject to $b^T x + c = 0$

(96)

Note that S and therefore also S^{-1} is symmetric. Using conventional calculus we have $\nabla f(x) = (S^{-1} + S^{-1^T})x - 2S^{-1}y = 2S^{-1}x - 2S^{-1}y$ and $\nabla g(x) = b^T$. Using the method of Lagrangian Multipliers [18] we have the system of equations (97)

$$2S^{-1}x - 2S^{-1}y + \lambda b^T = 0$$
$$b^T x + c = 0$$

(97)

This can be rewritten in block matrix form as shown in (98).

$$\begin{pmatrix} 2S^{-1} & b \\ b^T & 0 \end{pmatrix}\begin{pmatrix} x \\ \lambda \end{pmatrix} = \begin{pmatrix} 2S^{-1}y \\ -c \end{pmatrix}$$

(98)

---

[5]From the results of Section 5 the probability distributions will be Gaussian if the system dynamics are linear. However, it is well known [30] that MPC is not in general a linear controller. From an analytical point of view this is problematic. We assume that the nonlinearity introduced by the MPC is negligible.

The special structure of the left hand side matrix in (98) allows us to analytically compute the inverse [22] as shown in (99).

$$
\begin{pmatrix} 2S^{-1} & b \\ b^T & 0 \end{pmatrix}^{-1} = \begin{pmatrix} \frac{1}{2}S(I - \frac{bb^T S}{b^T Sb}) & \frac{Sb}{b^T Sb} \\ \frac{b^T S}{b^T Sb} & -\frac{2}{b^T Sb} \end{pmatrix}
\tag{99}
$$

To find the arguments which satisfy (97) we solve (100) which is equivalent to solving the system of linear equations in (98).

$$
\begin{pmatrix} \frac{1}{2}S(I - \frac{bb^T S}{b^T Sb}) & \frac{Sb}{b^T Sb} \\ \frac{b^T S}{b^T Sb} & -\frac{2}{b^T Sb} \end{pmatrix} \begin{pmatrix} 2S^{-1}y \\ -c \end{pmatrix} = \begin{pmatrix} S(I - \frac{bb^T S}{b^T Sb})S^{-1}y - c\frac{Sb}{b^T Sb} \\ 2(\frac{b^T S}{b^T Sb} + \frac{c}{b^T Sb}) \end{pmatrix}
\tag{100}
$$

Therefore, the arguments which minimise (96) are $x^* = S(I - \frac{bb^T S}{b^T Sb})S^{-1}y - c\frac{Sb}{b^T Sb}$. Substituting this into the objective function we have (101).

$$
\begin{aligned}
&(x^* - y)^T S^{-1} (x^* - y) \\
&= \left(S\left(I - \frac{bb^T S}{b^T Sb}\right) S^{-1}y - c\frac{Sb}{b^T Sb} - y \right)^T S^{-1} \left(S\left(I - \frac{bb^T S}{b^T Sb}\right) S^{-1}y - c\frac{Sb}{b^T Sb} - y \right) \\
&= \left(\frac{Sbb^T y}{b^T Sb} + c\frac{Sb}{b^T Sb} \right)^T S^{-1} \left(\frac{Sbb^T y}{b^T Sb} + c\frac{Sb}{b^T Sb} \right) \\
&= \frac{(Sb)^T}{b^T Sb} \left(b^T y + c\right)^T S^{-1} \frac{Sb}{b^T Sb} \left(b^T y + c\right) \\
&= \frac{b^T S}{b^T Sb} \left(b^T y + c\right)^T \frac{b}{b^T Sb} \left(b^T y + c\right) \\
&= \left(b^T y + c\right)^T \frac{b^T Sb}{(b^T Sb)^2} \left(b^T y + c\right) \\
&= \frac{(b^T y + c)^T (b^T y + c)}{b^T Sb} \\
&= \frac{(b^T y + c)^2}{b^T Sb}
\end{aligned}
\tag{101}
$$

We can conclude that the shortest squared Mahalanobis Distance between a point y and the constraints of (96) is $\frac{(b^T y + c)^2}{b^T Sb}$. $\qquad\square$

Theorem 7.7. Affine Chance Constraints

7.4 Discussion

Use the nonlinear underlying model.

References

[1] Y. Bar-Shalom, X.R. Li, and T. Kirubarajan. *Estimation with applications to tracking and navigation.* John Wiley and Sons, 2001.

[2] D. Barber. Expectation correction for smoothed inference in switching linear dynamical systems. *Journal of Machine Learning*, 7:2515–2540, 2006.

[3] D. Barber. *Bayesian Reasoning and Machine Learning.* Cambridge University Press, 2012.

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

[5] C.M. Bishop. *Pattern Recognition and Machine Learning.* Springer, 2006.

[6] L. Blackmore, O. Masahiro, A. Bektassov, and B.C. Williams. A probabilistic particle-control approximation of chance-constrained stochastic predictive control. *IEEE Transactions on Robotics*, 26, 2010.

[7] M. Cannon, B. Kouvaritakis, and X. Wu. Probabilistic constrained mpc for multiplicative and additive stochastic uncertainty. *IEEE Transactions on Automatic Control*, 54(7), 2009.

[8] A.L. Cervantes, O.E. Agamennoni, and J.L Figueroa. A nonlinear model predictive control system based on weiner piecewise linear models. *Journal of Process Control*, 13:655–666, 2003.

[9] R. Chen and J.S. Liu. Mixture kalman filters. *Journal of Royal Statistical Society*, 62(3):493–508, 2000.

[10] B.N. Datta. *Numerical Methods for Linear Control Systems - Design and Analysis.* Elsevier, 2004.

[11] M. Davidian. Applied longitudinal data analysis. North Carolina State University, 2005.

[12] N. Deo. *Graph Theory with Applications to Engineering and Computer Science.* Prentice-Hall, 1974.

[13] A. Doucet and A.M. Johansen. A tutorial on particle filtering and smoothing: fifteen years later. Technical report, The Institute of Statistical Mathematics, 2008.

[14] A.D. Doucet, N.J. Gordon, and V. Krishnamurthy. Particle filters for state estimation of jump markov linear systems. *IEEE Transactions on Signal Processing*, 49(3):613–624, March 2001.

[15] J. Du, C. Song, and P. Li. Modeling and control of a continuous stirred tank reactor based on a mixed logical dynamical model. *Chinese Journal of Chemical Engineering*, 15(4):533–538, 2007.

[16] The Economist. In praise of bayes. Article in Magazine, September 2000.

[17] H.C. Edwards and D.E. Penny. *Elementary Differential Equations*. Pearson, 6th edition edition, 2009.

[18] W. Forst and D. Hoffmann. *Optimisation - Theory and Practice*. Springer, 2010.

[19] O.R. Gonzalez and A.G. Kelkar. *Electrical Engineering Handbook*. Academic Press, 2005.

[20] N.J. Gordon, D.J. Salmond, and A.F.M. Smith. Novel approach to nonlinear/non-gaussian bayesian state estimation. *IEE Proceedings-F*, 140(2):107–113, 1993.

[21] K. Ito and K. Xiong. Gaussian filters for nonlinear filtering problems. *IEEE Transactions on Automatic Control*, 45(5):910–928, 2000.

[22] R. J. Jang and C.T. Sun. *Neuro-fuzzy and soft computing: a computational approach to learning and machine intelligence*. Prentice-Hall, 1996.

[23] D. Koller and N. Friedman. *Probabilistic Graphical Models*. MIT Press, 2009.

[24] K. B. Korb and A. E. Nicholson. *Bayesian Artificial Intelligence*. Series in Computer Science and Data Analysis. Chapman & Hall, first edition edition, 2004.

[25] M. Kvasnica, M. Herceg, L. Cirka, and M. Fikar. Model predictive control of a cstr: a hybrid modeling approach. *Chemical Papers*, 64(3):301–309, 2010.

[26] U.N. Lerner. *Hybrid Bayesian Networks for Reasoning about Complex Systems*. PhD thesis, Stanford Univesity, 2002.

[27] P. Li, M. Wendt, H. Arellano-Garcia, and G. Wozny. Optimal operation of distrillation processes under uncertain inflows accumulated in a feed tank. *American Institute of Chemical Engineers*, 2002.

[28] P. Li, M. Wendt, and G. Wozny. A probabilistically constrained model predictive controller. *Automatica*, 38:1171–1176, 2002.

[29] W.L. Luyben. *Process Modeling, Simulation and Control for Chemical Engineers*. McGraw-Hill, 2nd edition edition, 1990.

[30] J.M. Maciejowski. *Predictive Control with constraints*. Prentice-Hall, 2002.

[31] O. Masahiro. Joint chance-constrained model predictive control with probabilistic resolvability. *American Control Conference*, 2012.

[32] P. Mhaskar, N.H. El-Farra, and P.D. Christofides. Stabilization of nonlinear systems with state and control constraints using lyapunov-based predictive control. *Systems and Control Letters*, 55:650–659, 2006.

[33] K.P. Murphy. Switching kalman filters. Technical report, Compaq Cambridge Research Lab, 1998.

[34] K.P. Murphy. *Dynamic Bayesian Networks: Representation, Inference and Learning.* PhD thesis, University of California, Berkeley, 2002.

[35] K.P. Murphy. *Machine Learning: A Probabilistic Perspective.* MIT Press, 2012.

[36] T. Pan, S. Li, and W.J. Cai. Lazy learning based online identification and adaptive pid control: a case study for cstr process. *Industrial Engineering Chemical Research*, 46:472–480, 2007.

[37] J.B. Rawlings and D.Q. Mayne. *Model Predictive Control.* Nob Hill Publishing, 2009.

[38] A.T. Schwarm and Nikolaou. Chance constrained model predictive control. Technical report, University of Houston and Texas A&M University, 1999.

[39] S.J. Streicher, S.E. Wilken, and C. Sandrock. Eigenvector analysis for the ranking of control loop importance. *Computer Aided Chemical Engineering*, 33:835–840, 2014.

[40] D.H. van Hessem and O.H. Bosgra. Closed-loop stochastic dynamic process optimisation under input and state constraints. In *Proceedings of the American Control Conference*, 2002.

[41] D.H. van Hessem, C.W. Scherer, and O.H. Bosgra. Lmi-based closed-loop economic optimisation of stochastic process operation under state and input constraints. In *Proceedings of the 40th IEEE Conference on Decision and Control*, 2001.

[42] R.S. Wills. Google's pagerank: the math behind the search engine. Technical report, North Carolina State University, 2006.

[43] J. Yan and R.R. Bitmead. Model predictive control and state estimation: a network example. In *15th Triennial World Conference of IFAC*, 2002.

[44] J. Yan and R.R. Bitmead. Incorporating state estimation into model predictive control and its application to network traffic control. *Automatica*, 41:595–604, 2005.

[45] M.B. Yazdi and M.R. Jahed-Motlagh. Stabilization of a cstr with two arbitrarily switching modes using model state feedback linearisation. *Chemical Engineering Journal*, 155(3):838–843, 2009.