Electrical Engineering & Computer Science Faculty
Publications

Electrical Engineering & Computer Science
Department

8-1-2010

Kalman Filtering With State Constraints: A Survey of Linear and Nonlinear Algorithms

Daniel J. Simon
Cleveland State University, d.j.simon@csuohio.edu

Original Citation
Simon, D. (2010). Kalman filtering with state constraints: a survey of linear and nonlinear algorithms. IET Control Theory & Applications, 4(8), 1303-1318.

Repository Citation

KALMAN FILTERING WITH STATE CONSTRAINTS: A SURVEY OF LINEAR AND NONLINEAR ALGORITHMS

D. Simon, *Cleveland State University*

Department of Electrical and Computer Engineering, Cleveland State University, Stilwell Hall Room 332, 2121 Euclid Avenue, Cleveland, Ohio 44115, USA
E-mail: d.j.simon@csuohio.edu

Abstract

The Kalman filter is the minimum-variance state estimator for linear dynamic systems with Gaussian noise. Even if the noise is non-Gaussian, the Kalman filter is the best linear estimator. For nonlinear systems it is not possible, in general, to derive the optimal state estimator in closed form, but various modifications of the Kalman filter can be used to estimate the state. These modifications include the extended Kalman filter, the unscented Kalman filter, and the particle filter. Although the Kalman filter and its modifications are powerful tools for state estimation, we might have information about a system that the Kalman filter does not incorporate. For example, we may know that the states satisfy equality or inequality constraints. In this case we can modify the Kalman filter to exploit this additional information and get better filtering performance than the Kalman filter provides. This paper provides an overview of various ways to incorporate state constraints in the Kalman filter and its nonlinear modifications. If both the system and state constraints are linear, then all of these different approaches result in the same state estimate, which is the optimal constrained linear state estimate. If either the system or constraints are nonlinear, then constrained filtering is, in general, not optimal, and different approaches give different results.

1   Introduction

The Kalman filter is the minimum-variance state estimator for linear dynamic systems with Gaussian noise [1]. In addition, the Kalman filter is the minimum-variance linear state estimator for linear dynamic systems with non-Gaussian noise [2]. For nonlinear systems it is not possible, in general, to implement the optimal state estimator in closed form, but various modifications of the Kalman filter can be used to estimate the state. These modifications include the extended Kalman filter [2], the unscented Kalman filter [3], and the particle filter [4].

Although the Kalman filter and its modifications are powerful tools for state estimation, we might have information about a system that the Kalman filter does not incorporate. For example, we may know that the states satisfy equality or inequality constraints. In this case we can modify the Kalman filter to exploit this additional information and get better filtering performance than the Kalman filter provides.

An initial consideration leads us to believe that the incorporation of constraints cannot improve the performance of the Kalman filter. After all, since the Kalman filter is minimum variance, it should not be possible to improve it. However, there are two reasons that the Kalman filter can indeed be improved. First, if the system is nonlinear, the Kalman filter variations are only approximately minimum variance, so it is not surprising that improvements can be seen by incorporating state constraints in the filter. Second, even if the system is linear, if there are additional constraints beyond those explicitly given in the system model, then the complete system description is different than that assumed by the standard Kalman filter equations, and a modification of the Kalman filter may result in improved performance.

We see many examples of state-constrained systems in engineering applications. Some of these examples include camera tracking [5], fault diagnosis [6], chemical processes [7], vision-based systems [8], target tracking [9, 10], biomedical systems [11], robotics [12], navigation [13], and

others [14]. This paper presents a survey of how state constraints can be incorporated into the Kalman filter and its nonlinear modifications. We discuss linear and nonlinear systems, linear and nonlinear state constraints, and equality and inequality state constraints.

Section 2 considers linear systems and linear constraints. The various ways of enforcing linear constraints in the linear Kalman filter include model reduction [15], perfect measurements [8–10], estimate projection [13, 16], gain projection [17, 18], probability density function (PDF) truncation [2, 19, 20], and system projection [21]. Under certain conditions, all these approaches result in the same state estimate. We also briefly discuss inequality constraints and soft constraints. We present an example illustrating the conditions under which these approaches are equivalent, and conditions under which these approaches differ.

Section 3 considers systems that are nonlinear or that have nonlinear constraints. The approaches that can be used in these cases include second-order expansion of the constraints [22], the smoothly constrained Kalman filter [23], moving horizon estimation [24–26], various modifications of the unscented Kalman filter [2, 3], interior point approaches [27], and particle filters [2, 28]. We present an example showing that moving horizon estimation performs the best relative to estimation error, but this performance comes at a high computational expense.

2 The Kalman filter

Consider the system model

$$x_{k+1} = Fx_k + w_k, \tag{1}$$

$$y_k = Hx_k + v_k, \tag{2}$$

where k is the time step, x_k is the state, y_k is the measurement, w_k and v_k are the zero-mean process noise and measurement noise with covariances Q and R respectively, and F and H are the state transition and measurement matrices. The Kalman filter was independently invented in the 1950's by several different researchers and is named after Rudolph Kalman [29]. The Kalman filter equations are given as [2]

$$P_k^- = FP_{k-1}^+F^T + Q, \tag{3}$$

$$K_k = P_k^-H^T(HP_k^-H^T + R)^{-1}, \tag{4}$$

$$\hat{x}_k^- = F\hat{x}_{k-1}^+, \tag{5}$$

$$\hat{x}_k^+ = \hat{x}_k^- + K_k(y_k - H\hat{x}_k^-), \tag{6}$$

$$P_k^+ = (I - K_kH)P_k^-, \tag{7}$$

for $k = 1, 2, \ldots$, where I is the identity matrix. $\hat{x}_k^-$ is the *a priori* estimate of the state x_k given measurements up to and including time $k - 1$. $\hat{x}_k^+$ is the *a posteriori* estimate of the state x_k given measurements up to and including time

k. K_k is the Kalman gain, P_k^- is the covariance of the *a priori* estimation error $x_k - \hat{x}_k^-$, and P_k^+ is the covariance of the *a posteriori* estimation error $x_k - \hat{x}_k^+$. The Kalman filter is initialised with

$$\hat{x}_0^+ = E(x_0), \tag{8}$$

$$P_0^+ = E[(x_0 - \hat{x}_0^+)(x_0 - \hat{x}_0^+)^T], \tag{9}$$

where $E(\cdot)$ is the expectation operator.

When the noise sequences $\{w_k\}$ and $\{v_k\}$ are Gaussian, uncorrelated, and white, the Kalman filter is the minimum-variance filter and minimises the trace of the estimation error covariance at each time step. When $\{w_k\}$ and $\{v_k\}$ are non-Gaussian, the Kalman filter is the minimum-variance linear filter, although there might be nonlinear filters that perform better [30]. When $\{w_k\}$ and $\{v_k\}$ are correlated or colored, (3)–(7) can be modified to obtain the minimum-variance filter [2].

Now suppose that our system satisfies the equality constraints

$$Dx_k = d, \tag{10}$$

or the inequality constraints

$$Dx_k \leq d, \tag{11}$$

where D is a known matrix and d is a known vector. In this case we might want to find a state estimate $\hat{x}_k$ that satisfies the constraints

$$D\hat{x}_k = d, \tag{12}$$

or

$$D\hat{x}_k \leq d. \tag{13}$$

In the following sections we discuss several ways to modify the Kalman filter to incorporate these linear equality and inequality constraints.

2.1 Model reduction

Equality constraints in the form of (10) can be addressed by reducing the system model parameterization [15]. As an example, consider the system

$$x_{k+1} = \begin{bmatrix} 1 & 2 & 3 \\ 3 & 2 & 1 \\ 4 & -2 & 2 \end{bmatrix} x_k + w_k, \tag{14}$$

$$y_k = \begin{bmatrix} 2 & 4 & 5 \end{bmatrix} x_k + v_k. \tag{15}$$

Suppose that we also have the constraint

$$\begin{bmatrix} 1 & 0 & 1 \end{bmatrix} x_k = 0. \tag{16}$$

If we make the substitution $x_k(3) = -x_k(1)$ in (14) and (15), we obtain

$$x_{k+1}(1) = -2x_k(1) + 2x_k(2) + w_{1k}, \qquad (17)$$

$$x_{k+1}(2) = 2x_k(1) + 2x_k(2) + w_{2k}, \qquad (18)$$

$$y_k = -3x_k(1) + 4x_k(2) + v_k. \qquad (19)$$

(17)–(19) can be written as

$$x_{k+1} = \begin{bmatrix} -2 & 2 \\ 2 & 2 \end{bmatrix} x_k + \begin{bmatrix} w_{1k} \\ w_{2k} \end{bmatrix}, \qquad (20)$$

$$y_k = \begin{bmatrix} -3 & 4 \end{bmatrix} x_k + v_k. \qquad (21)$$

This example shows how to reduce an equality-constrained filtering problem to an equivalent but unconstrained filtering problem. The Kalman filter for this unconstrained system is the optimal linear estimator, and thus it is also the optimal linear estimator for the original constrained system. The dimension of the reduced model is lower than that of the original model, which reduces the computational effort of the Kalman filter. One disadvantage of this approach is that the physical meaning of the state variables may be lost. Also this approach cannot be directly used for inequality constraints in the form of (11).

2.2 Perfect measurements

State equality constraints can be treated as perfect measurements with zero measurement noise [8–10]. If the constraints are given by (10), where D is an $s \times n$ matrix with $s < n$, then we can augment (2) with s perfect measurements of the state.

$$\begin{bmatrix} y_k \\ d \end{bmatrix} = \begin{bmatrix} H \\ D \end{bmatrix} x_k + \begin{bmatrix} v_k \\ 0 \end{bmatrix}. \qquad (22)$$

The state equation (1) is not changed, but the measurement equation is augmented. The fact that the last s components of the measurement equation are noise free means that the *a posteriori* Kalman filter estimate of the state is consistent with these s measurements [31]. That is, the Kalman filter estimate satisfies $D\hat{x}_k^+ = d$. This approach is mathematically identical to the model reduction approach.

Note that the new measurement noise covariance is singular. A singular noise covariance does not, in general, present theoretical problems [23]. However, in practice a singular noise covariance increases the possibility of numerical problems such as ill conditioning in the covariance of the error estimate [32, p. 249], [33, p. 365]. Also the use of perfect measurements is not directly applicable to inequality constraints in the form of (11).

2.3 Estimate projection

Another approach to constrained filtering is to project the unconstrained estimate $\hat{x}_k^+$ of the Kalman filter onto the constraint surface [13, 16]. The constrained estimate can therefore be written as

$$\tilde{x}_k^+ = \text{argmin}_x (x - \hat{x}_k^+)^T W (x - \hat{x}_k^+), \qquad (23)$$

such that

$$Dx = d, \qquad (24)$$

where W is a positive-definite weighting matrix. The solution to this problem is

$$\tilde{x}_k^+ = \hat{x}_k^+ - W^{-1}D^T(DW^{-1}D^T)^{-1}(D\hat{x}_k^+ - d). \qquad (25)$$

If the process and measurement noises are Gaussian and we set $W = (P_k^+)^{-1}$ we obtain the maximum probability estimate of the state subject to state constraints. If we set $W = I$ we obtain the least squares estimate of the state subject to state constraints. This approach is similar to that used in [34] for input signal estimation. See [2, p. 218] for a graphical interpretation of the projection approach to constrained filtering.

It is shown in [13, 16] that the constrained state estimate of (25) is unbiased. That is,

$$E(\tilde{x}_k^+) = E(x_k). \qquad (26)$$

Setting $W = (P_k^+)^{-1}$ results in the minimum variance filter. That is, if $W = (P_k^+)^{-1}$ then

$$\text{Cov}(x_k - \tilde{x}_k^+) \leq \text{Cov}(x_k - \hat{x}_k^+), \qquad (27)$$

for all $\hat{x}_k^+$. Setting $W = I$ results in a constrained estimate that is closer to the true state than the unconstrained estimate at each time step. That is, if $W = I$ then

$$\|x_k - \tilde{x}_k^+\|_2 \leq \|x_k - \hat{x}_k^+\|_2, \qquad (28)$$

for all k.

(25) was obtained for $W = (P_k^+)^{-1}$ in [14] in a different form along with some additional properties and generalisations. It is assumed in [13, 16] that the constrained *a priori* estimate is based on the unconstrained estimate so that the constrained filter is

$$\hat{x}_k^- = F\hat{x}_{k-1}^+, \qquad (29)$$

$$\hat{x}_k^+ = \hat{x}_k^- + K_k(y_k - H\hat{x}_k^-), \qquad (30)$$

$$\tilde{x}_k^+ = \hat{x}_k^+ - P_k^+ D^T(DP_k^+ D^T)^{-1}(D\hat{x}_k^+ - d). \qquad (31)$$

If the constrained *a priori* estimate is based on the

constrained estimate [14] then the constrained filter is

$$\hat{x}_k^- = F\tilde{x}_{k-1}^+, \tag{32}$$

$$\hat{x}_k^+ = \hat{x}_k^- + K_k(y_k - H\hat{x}_k^-), \tag{33}$$

$$\tilde{x}_k^+ = \hat{x}_k^+ - P_k^+ D^T (D P_k^+ D^T)^{-1}(D\hat{x}_k^+ - d). \tag{34}$$

It can be inductively shown that (29)–(31) and (32)–(34) result in the same state estimates when $\hat{x}_0^+$ in (29) is equal to $\tilde{x}_0^+$ in (32). It can also be shown that these constrained estimates are the same as those obtained with the perfect measurement approach [14, 35].

2.4 Estimate projection with inequality constraints

The estimate projection approach to constrained filtering has the advantage that it can be extended to inequality constraints in the form of (13). If we have the constraints $D\hat{x}_k \le d$, then a constrained estimate can be obtained by modifying (23)–(24) and solving the problem

$$\tilde{x}_k^+ = \mathrm{argmin}_x (x - \hat{x}_k^+)^T W (x - \hat{x}_k^+), \tag{35}$$

such that

$$Dx \le d. \tag{36}$$

This problem is a quadratic programming problem [36, 37]. Various approaches can be used to solve quadratic programming problems, including interior point approaches and active set methods [38]. An active set method uses the fact that it is only those constraints that are active at the solution of the problem that are significant in the optimality conditions. Suppose that we have s inequality constraints, and q of the s inequality constraints are active at the solution of (35)–(36). Denote by $\hat{D}$ the q rows of D that correspond to the active constraints, and denote by $\hat{d}$ the q components of d that correspond to the active constraints. If the set of active constraints is known *a priori* then the solution of (35)–(36) is also a solution of the equality-constrained problem

$$\tilde{x}_k^+ = \mathrm{argmin}_x (x - \hat{x}_k^+)^T W (x - \hat{x}_k^+), \tag{37}$$

such that

$$\hat{D}x = \hat{d}. \tag{38}$$

The inequality-constrained problem of (35)–(36) is equivalent to the equality-constrained problem of (37)–(38). Therefore all of the properties of the equality-constrained state estimate also apply to the inequality-constrained state estimate.

2.5 Gain projection

The standard Kalman filter can be derived by solving the problem [2]

$$K_k = \mathrm{argmin}_K \mathrm{Trace}[(I - KH)P_k^-(I - KH)^T + KRK]. \tag{39}$$

The solution to this problem gives the optimal Kalman gain

$$S_k = H P_k^- H^T + R, \tag{40}$$

$$K_k = P_k^- H^T S_k^{-1}, \tag{41}$$

and the state estimate measurement update is

$$r_k = y_k - H\hat{x}_k^-, \tag{42}$$

$$\hat{x}_k^+ = \hat{x}_k^- + K_k r_k. \tag{43}$$

If the constraint $D\hat{x}_k^+ = d$ is added to the problem, then the minimization problem of (39) can be written as

$$\tilde{K}_k = \mathrm{argmin}_K \mathrm{Trace}[(I - KH)P_k^-(I - KH)^T + KRK], \tag{44}$$

such that

$$D\hat{x}_k^+ = d. \tag{45}$$

The solution to this constrained problem is [17]

$$\tilde{K}_k = K_k - D^T(DD^T)^{-1}(D\hat{x}_k^+ - d)(r_k^T S_k^{-1} r_k)^{-1} r_k^T S_k^{-1}. \tag{46}$$

When this value for $\tilde{K}_k$ is used in place of K_k in (43), the result is the constrained state estimate

$$\tilde{x}_k^+ = \hat{x}_k^+ - D^T(DD^T)^{-1}(D\hat{x}_k^+ - d). \tag{47}$$

This estimate is the same as that given in (25) with $W = I$.

Gain projection has been applied to inequality constraints in [18]. If the *a priori* estimate $\hat{x}_k^-$ satisfies the constraints and the unconstrained *a posteriori* estimate $\hat{x}_k^+$ does not satisfy them, then $\hat{x}_k^-$ can be projected in the direction of $\hat{x}_k^+$ until it reaches the constraint boundary. This effectively gives a modified Kalman gain $K_k^{(m)} = \beta K_k$, where $\beta \in (0, 1)$ and K_k is the standard unconstrained Kalman gain.

2.6 Probability density function truncation

In the PDF truncation approach, we take the PDF of the state estimate that is computed by the Kalman filter, assuming that it is Gaussian, and truncate the PDF at the constraint edges. The constrained state estimate is equal to the mean of the truncated PDF [2, 19, 20]. This approach is designed for inequality constraints on the state although

it can also be applied to equality constraints with a simple modification. See [2, p. 222] for a graphical illustration of how this method works.

This method is complicated when the state dimension is more than one. In that case the state estimate is normalized so that its components are statistically independent of each other. Then the normalized constraints are applied one at a time. After all the constraints are applied, the normalization process is reversed to obtain the constrained state estimate. Details of the algorithm are given in [2, 20].

2.7 System projection

State constraints imply that the process noise is also constrained. This realisation leads to a modification of the initial estimation error covariance and the process noise covariance, after which the standard Kalman filter equations are implemented [21]. Given the constrained system

$$x_{k+1} = Fx_k + w_k, \tag{48}$$

$$Dx_k = d, \tag{49}$$

it is reasonable to suppose that the noise-free system also satisfies the constraints. That is, $DFx_k = 0$. But this result means that $Dw_k = 0$. If these equations are not satisfied, then the noise w_k is correlated with the state x_k, which violates typical assumptions on the system characteristics. If $Dw_k = 0$ then

$$Dw_k w_k^T D^T = 0, \tag{50}$$

$$E(Dw_k w_k^T D^T) = 0, \tag{51}$$

$$DQD^T = 0. \tag{52}$$

This equation means that Q must be singular, assuming that D has full row rank. As a simple example consider the three-state system given in (14)–(15). From (14) we have

$$x_{1,k+1} + x_{3,k+1} = 5x_{1k} + 5x_{3k} + w_{1k} + w_{3k}. \tag{53}$$

Combining this equation with (16) gives

$$w_{1k} + w_{3k} = 0, \tag{54}$$

which means the covariance matrix Q must be singular for this constrained system to be consistent. We must have $Dw_k = 0$, which in turn implies (52).

If the process noise covariance Q does not satisfy (52) then it can be projected onto a modified covariance $\tilde{Q}$ that does satisfy the constraint to make the system consistent. $\tilde{Q}$ then replaces Q in the Kalman filter. The formulation of $\tilde{Q}$ is accomplished as follows [21]. First we find the singular value decomposition of D^T.

$$D^T = USV^T = \begin{bmatrix} U_1 & U_2 \end{bmatrix} \begin{bmatrix} S_r & 0 \end{bmatrix} V^T, \tag{55}$$

where S_r is an $r \times r$ matrix, and r is the rank of D. Next we compute $N = U_2 U_2^T$, which is the orthogonal projector onto the null space of D. Next we compute $\tilde{Q} = NQN$. This value for $\tilde{Q}$ ensures that

$$D\tilde{Q}D^T = (VS^T U^T)(U_2 U_2^T Q U_2 U_2^T)(USV^T) = 0, \tag{56}$$

and thus (52) is satisfied. Similarly the initial estimation error covariance can be modified as $\tilde{P}_0^+ = NP_0^+ N$ in order to be consistent with the state constraints. It is shown in [21] that the estimation error covariance obtained by this method is less than or equal to that obtained by the estimate projection method. The reason for this conclusion is that $\tilde{Q}$ is assumed to be the true process noise covariance, so that the system projection method gives the optimal state estimate, just as the standard Kalman filter gives the optimal state estimate for an unconstrained system. But the standard Kalman filter, and estimate projection methods based on it, might use an incorrect covariance Q. However, if Q satisfies $DQD^T = 0$ then the standard Kalman filter estimate satisfies the state constraint $D\hat{x}_k^+ = 0$, and the system projection filter, the estimate projection filter, and the standard Kalman filter are all identical.

2.8 Soft constraints

Soft constraints, as opposed to hard constraints, are constraints that are only required to be approximately satisfied rather than exactly satisfied. We might want to implement soft constraints in cases where the constraints are heuristic rather than rigorous, or in cases where the constraint function has some uncertainty or fuzziness. For example, suppose we have a vehicle navigation system with two states: $x(1)$, which is north position, and $x(2)$, which is east position. We know that the vehicle is on a straight road such that $x(1) = mx(2) + b$ for known constants m and b. But the road also has an unknown nonzero width, so the state constraint can be written as $x(1) \approx mx(2) + b$. In this case we have an approximate equality constraint, which is referred to in the literature as a soft constraint. It can be argued that estimators for most practical engineering systems should be implemented with soft constraints rather than hard constraints.

Soft constraints can be implemented in Kalman filters in various ways. First, the perfect measurement approach can be extended to soft constraints by adding small nonzero measurement noise to the perfect measurements [9, 10, 39, 40]. Second, soft constraints can be implemented by adding a regularization term to the standard Kalman filter [6]. Third, soft constraints can be enforced by projecting the unconstrained estimates in the direction of the constraints rather than exactly onto the constraint surface [41].

Example 1: Consider a navigation problem. The first two state components are the north and east positions of a land

vehicle, and the last two components are the north and east velocities. The velocity of the vehicle is in the direction of θ, an angle measured clockwise from due east. A position-measuring device provides a noisy measurement of the vehicle's north and east positions. Equations for this system can be written as

$$x_{k+1} = \begin{bmatrix} 1 & 0 & T & 0 \\ 0 & 1 & 0 & T \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} x_k + \begin{bmatrix} 0 \\ 0 \\ T\sin\theta \\ T\cos\theta \end{bmatrix} u_k + w_k, \quad (57)$$

$$y_k = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix} x_k + v_k, \quad (58)$$

where T is the discretisation step size and u_k is an acceleration input. The covariances of the process and measurement noise are $Q = \mathrm{diag}(4, \quad 4, \quad 1, \quad 1)$ and $R = \mathrm{diag}(900, \quad 900)$. The initial estimation error covariance is $P_0^+ = \mathrm{diag}(900, 900, 4, 4)$. If we know that the vehicle is on a road with a heading of θ then we have

$$\tan\theta = x(1)/x(2) = x(3)/x(4). \quad (59)$$

We can write these constraints in the form $D_i x_k = 0$ using one of two D_i matrices.

$$D_1 = \begin{bmatrix} 1 & -\tan\theta & 0 & 0 \\ 0 & 0 & 1 & -\tan\theta \end{bmatrix}, \quad (60)$$

$$D_2 = \begin{bmatrix} 0 & 0 & 1 & -\tan\theta \end{bmatrix}. \quad (61)$$

D_1 directly constrains both velocity and position. D_2 relies on the fact that velocity determines position, which means that a velocity constraint implicitly constrains position. Note that we cannot use $D = \begin{bmatrix} 1 & -\tan\theta & 0 & 0 \end{bmatrix}$. If we did then position would be constrained but velocity would not be constrained. But it is velocity that determines position through the system equations, therefore this value of D is not consistent with the

state equations. In particular it violates the $DF = D$ condition of [14].

At this point we can take several approaches to state estimation. For example, we can use Q and P_0^+ to run the standard unconstrained Kalman filter and ignore the constraints, or run the perfect measurement filter, or project the unconstrained estimate onto the constraint surface, or use the PDF truncation method, or use constrained moving horizon estimation (MHE). MHE is discussed later in this paper since it is a general nonlinear estimator. Alternatively we can use the projected $\tilde{Q}$ and $\tilde{P}_0^+$ and then run the standard Kalman filter. Since $\tilde{Q}$ and $\tilde{P}_0^+$ are consistent with the constraints, the state estimate satisfies the constraints for all time if the initial estimate $\hat{x}_0^+$ satisfies the constraints. This approach is the system projection approach. Note that neither the perfect measurement filter, the estimate projection filter, the PDF truncation filter, nor MHE, changes the estimate in this case, since the unconstrained estimate is implicitly constrained by means of system projection. In addition to all these options, we can choose to use either the D_1 or D_2 matrix of (60)–(61) to constrain the system.

MATLAB® software was written to implement these constrained filtering algorithms on a 150-s simulation with a 3-s simulation step size [42]. We used the initial state $x_0 = \begin{bmatrix} 0 & 0 & 10\tan\theta & 10 \end{bmatrix}^T$ and perfect initial state estimates. Table 1 shows the RMS state estimation errors averaged for the two position states, and the RMS constraint error. Each RMS value shown is averaged over 100 Monte Carlo simulations. Table 1 shows that all of the constrained filters have constraint errors that are exactly zero. All of the constrained filters perform identically when D_1 is used as the constraint matrix. However, when D_2 is used as the constraint matrix, then the perfect measurement and system projection methods perform the best.

Table 1 Filter results for the linear vehicle navigation problem

Filter type	RMS estimation error (D_1, D_2)	RMS constraint error (D_1, D_2)
unconstrained	23.7, 23.7	31.7, 2.1
perfect measurement	17.3, 19.2	0, 0
estimate projection	17.3, 21.4	0, 0
MHE, horizon size 2	17.3, 20.3	0, 0
MHE, horizon size 4	17.3, 19.4	0, 0
system projection	17.3, 19.2	0, 0
PDF truncation	17.3, 21.4	0, 0

Two numbers in each cell indicate the errors that are obtained using the D_1 and D_2 constraints respectively. The numbers shown are RMS errors averaged over 100 Monte Carlo simulations

3 Nonlinear constraints

Sometimes state constraints are nonlinear. Instead of $Dx_k = d$ we have

$$g(x_k) = h. \tag{62}$$

We can perform a Taylor series expansion of the constraint equation around $\hat{x}_k^-$ to obtain

$$g(x_k) \approx g(\hat{x}_k^-) + g'(\hat{x}_k^-)(x_k - \hat{x}_k^-)$$
$$+ \frac{1}{2} \sum_{i=1}^{s} e_i (x_k - \hat{x}_k^-)^T g_i''(\hat{x}_k^-)(x_k - \hat{x}_k^-), \tag{63}$$

where s is the dimension of $g(x)$, e_i is the the ith natural basis vector in $\mathcal{R}^s$, and the entry in the pth row and qth column of the $n \times n$ matrix $g_i''(x)$ is given by

$$[g_i''(x)]_{pq} = \frac{\partial^2 g_i(x)}{\partial x_p \partial x_q}. \tag{64}$$

Neglecting the second-order term gives [8, 9, 13, 31]

$$g'(\hat{x}_k^-)x_k = h - g(\hat{x}_k^-) + g'(\hat{x}_k^-)\hat{x}_k^-. \tag{65}$$

This equation is equivalent to the linear constraint $Dx_k = d$ if

$$D = g'(\hat{x}_k^-), \tag{66}$$

$$d = h - g(\hat{x}_k^-) + g'(\hat{x}_k^-)\hat{x}_k^-. \tag{67}$$

Therefore all of the methods presented in Section 2 can be used with nonlinear constraints after the constraints are linearized. Sometimes, though, we can do better than simple linearization, as discussed in the following sections.

3.1 Second-order expansion

If we keep the second-order term in (63) then the constrained estimation problem can be approximately written as

$$\tilde{x}_k^+ = \text{argmin}_x (x - \hat{x}_k^+)^T W (x - \hat{x}_k^+), \tag{68}$$

such that

$$x^T M_i x + 2 m_i^T x + \mu_i = 0 \quad (i = 1, \dots, s), \tag{69}$$

where W is a weighting matrix, and M_i, m_i, and μ_i are obtained from (63) as

$$M_i = g_i''(\hat{x}_k^-)/2, \tag{70}$$

$$m_i = (g_i'(\hat{x}_k^-) - (\hat{x}_k^-)^T g_i''(\hat{x}_k^-))^T / 2, \tag{71}$$

$$\mu_i = g_i(\hat{x}_k^-) - g_i'(\hat{x}_k^-)\hat{x}_k^- + (\hat{x}_k^-)^T M_i \hat{x}_k^- - h_i. \tag{72}$$

This idea is similar to the way that the extended Kalman filter (EKF), which relies on linearization of the system and measurement equations, can be improved by retaining second-order terms to obtain the second-order EKF [2]. The optimization problem given in (68)-(69) can be solved with a numerical method. A Lagrange multiplier method for solving this problem is given in [22] for $s = 1$ and M positive definite.

3.2 The smoothly constrained Kalman filter

Another approach to handling nonlinear equality constraints is the smoothly constrained Kalman filter (SCKF) [23]. This approach starts with the idea that nonlinear constraints can be handled by linearizing them and then implementing them as perfect measurements. However, the resulting estimate only approximately satisfies the nonlinear constraint. If the constraint linearization is instead applied multiple times at each measurement time then the resulting estimate is expected to get closer to constraint satisfaction with each iteration. This idea is similar to the iterated Kalman filter for unconstrained estimation [2]. In the iterated Kalman filter the nonlinear system is repeatedly linearized at each measurement time. In the SCKF the nonlinear constraints are linearized at each time step and are repeatedly applied as measurements with increasing degrees of certainty. This idea is motivated by realizing that, for example, incorporating a measurement with a variance of 1 is equivalent to incorporating that same measurement 10 times, each with a variance of 10. Application of a scalar nonlinear constraint $g(x) = h$ by means of the SCKF is performed by the following algorithm, which is executed after each measurement update.

1. Initialize i, the number of constraint applications, to 1. Initialize $\hat{x}$ to $\hat{x}_k^+$, and P to P_k^+.

2. Set $R_0' = \alpha G P G^T$, where the $1 \times n$ Jacobian $G = g'(\hat{x})$. R_0' is the initial variance with which the constraint is incorporated into the state estimate as a measurement. Note that $G P G^T$ is the approximate linearized variance of $g(x)$, therefore R_0' is the fraction of this variance that is used to incorporate the constraint as a measurement. α is a tuning parameter, typically between 0.01 and 0.1.

3. Set $R_i' = R_0' \exp(-i)$. This equation is used to gradually decrease the measurement variance that is used to apply the constraint.

4. Set $S_i = \max_j (G_j P_{jj} G_j)/(G P G^T)$. S_i is a normalized version of the information that is associated with the constraint. When S_i exceeds the threshold $S_{\max}$ then the iteration is terminated. A typical value of $S_{\max}$ is 100. The iteration can also be terminated after a predetermined number of constraint applications $i_{\max}$, since a convergence proof for the SCKF does not yet exist. After the iteration terminates, set $\hat{x}_k^+ = \hat{x}$ and $P_k^+ = P$.

5. Incorporate the constraint as a measurement using

$$K = PG^T(GPG^T + R_i')^{-1}, \qquad (73)$$

$$\hat{x} = \hat{x} + K(h - g(\hat{x})), \qquad (74)$$

$$P = P(I - KG). \qquad (75)$$

These equations are the standard Kalman filter equations for a measurement update, but the measurement that we are incorporating is the not-quite-perfect measurement of the constraint.

6. Compute the updated Jacobian $G = g'(\hat{x})$. Increment i by one and go to step 3 to continue the iteration.

The above algorithm executes once for each inequality constraint at each time step.

3.3 Moving horizon estimation

Moving horizon estimation (MHE), first suggested in [43], is based on the fact that the Kalman filter solves the optimization problem

$$\{\hat{x}_k^+\} = \mathrm{argmin}_{\{x_k\}} \|x_0 - \hat{x}_0\|_{I_0^+}^2 + \sum_{k=1}^{N} \|y_k - Hx_k\|_{R^{-1}}^2$$

$$+ \sum_{k=0}^{N-1} \|x_{k+1} - Fx_k\|_{Q^{-1}}^2, \qquad (76)$$

where N is the number of measurements that are available [24–26]. $\{\hat{x}_k^+\}$ is the sequence of estimates $\hat{x}_0^+, \ldots, \hat{x}_N^+$, and $I_0^+ = (P_0^+)^{-1}$. (76) is a quadratic programming problem. The $\{\hat{x}_k^+\}$ sequence that solves this problem gives the optimal smoothed estimate of the state given the measurements $y_1, \ldots, y_N$.

The above discussion motivates a similar method for general nonlinear constrained estimation [25, 26, 44]. Given

$$x_{k+1} = f(x_k) + w_k, \qquad (77)$$

$$y_k = h(x_k) + v_k, \qquad (78)$$

$$g(x_k) = 0, \qquad (79)$$

solve the optimization problem

$$\min_{\{x_k\}} \|x_0 - \hat{x}_0\|_{I_0^+}^2 + \sum_{k=1}^{N} \|y_k - h(x_k)\|_{R^{-1}}^2 + \sum_{k=0}^{N-1} \|x_{k+1} - f(x_k)\|_{Q^{-1}}^2,$$

$$(80)$$

such that

$$g(\{x_k\}) = 0, \qquad (81)$$

where by an abuse of notation we use $g(\{x_k\})$ to mean $g(x_k)$ for $k = 1, \ldots, N$. This constrained nonlinear optimization problem can be solved by various methods [36, 45, 46], therefore all of the theory that applies to the particular optimization algorithm that is used also applies to constrained MHE. The difficulty is the fact that the dimension of the problem increases with time. With each measurement that is obtained, the number of independent variables increases by n, where n is the number of state variables.

MHE therefore limits the time span of the problem to decrease the computational effort. The MHE problem can be written as

$$\min_{\{x_k\}} \|x_M - \hat{x}_M^+\|_{I_M^+}^2 + \sum_{k=M+1}^{N} \|y_k - h(x_k)\|_{R^{-1}}^2$$

$$+ \sum_{k=M}^{N-1} \|x_{k+1} - f(x_k)\|_{Q^{-1}}^2, \qquad (82)$$

such that

$$g(\{x_k\}) = 0, \qquad (83)$$

where $\{x_k\}$ is the set $\{x_M, \ldots, x_N\}$, and $N - M + 1$ is the horizon size. The dimension of this problem is $n(N - M + 1)$. The horizon size is chosen to give a tradeoff between estimation accuracy and computational effort. The information matrix I_M^+ is the inverse of P_M^+. The approximate estimation error covariance P_M^+ is obtained from the standard EKF recursion [2].

$$F_{k-1} = \left.\frac{\partial f}{\partial x}\right|_{\hat{x}_{k-1}^+}, \qquad (84)$$

$$H_k = \left.\frac{\partial h}{\partial x}\right|_{\hat{x}_k^-}, \qquad (85)$$

$$P_k^- = F_{k-1}P_{k-1}^+F_{k-1}^T + Q, \qquad (86)$$

$$K_k = P_k^- H_k^T (H_k P_k^- H_k^T + R)^{-1}, \qquad (87)$$

$$P_k^+ = (I - K_k H_k)P_k^-. \qquad (88)$$

Some stability results related to MHE are given in [47]. MHE is attractive in the generality of its formulation, but this generality results in large computational effort compared to the various constrained EFKs and unscented Kalman filters (UKFs), even for small horizons.

Another difficulty with MHE is its assumption of an invertible P_0^+ in (76) and (80), and an invertible P_M^+ in (82). The estimation error covariance for a constrained system is usually singular [13]. We can get around this by using the covariance of the unconstrained filter as shown in (88), but this makes MHE suboptimal even for linear systems.

Another way to deal with this is to use the singular constrained covariance shown in [13] and reduce it to a diagonal form [48, pp. 30−31]. This results in a corresponding transformation of the state estimate. Some of the transformed state estimates will have a zero variance, which means that those estimates will not change from one time step to the next in (82). This gives a more optimal implementation of MHE, but at the cost of additional complexity.

Recursive nonlinear dynamic data reconciliation and combined predictor-corrector optimization [7] are other approaches to constrained state estimation that are similar to MHE. These methods are essentially MHE with a horizon size of one. However the ultimate goal of these methods is data reconciliation (that is, output estimation) rather than state estimation, and they also include parameter estimation.

3.4 Unscented Kalman filtering

The unscented Kalman filter (UKF) is a filter for nonlinear systems that is based on two fundamental principles [2, 3]. First, although it is difficult to perform a nonlinear transformation of a PDF, it is easy to perform a nonlinear transformation of a vector. Second, it is not difficult to find a set of vectors in state space whose sample PDF approximates a given PDF. The UKF operates by producing a set of vectors called sigma points. The UKF uses between $n + 1$ and $2n + 1$ sigma points, where n is the dimension of the state. The sigma points are transformed and combined in a special way in order to obtain an estimate of the state and an estimate of the covariance of the state estimation error. Constraints can be incorporated into the UKF by treating the constraints as perfect measurements, which can be done in various ways as discussed below.

One possibility is to base the *a priori* state estimate on the unconstrained UKF *a posteriori* state estimate from the previous time step [14, 49]. In this case the standard unconstrained UKF runs independently of the constrained UKF. At each measurement time the state estimate of the unconstrained UKF is combined with the constraints, which are treated as perfect measurements, to obtain a constrained *a posteriori* UKF estimate. This filter is referred to as the projected UKF (PUKF) and is analogous to (29)−(31) for linear systems and constraints. Note that nonlinear constraints can be incorporated as perfect measurements in various ways, such as linearization, second-order expansion [22], unscented transformation [5], or the SCKF, which is an open research problem.

Another approach is to base the *a priori* state estimate on the constrained UKF *a posteriori* state estimate from the previous time step [14]. At each measurement time the state estimate of the unconstrained UKF is combined with the constraints, which are treated as perfect measurements,

to obtain a constrained *a posteriori* UKF estimate. This constrained *a posteriori* estimate is then used as the initial condition for the next time update. This filter is referred to as the equality constrained UKF (ECUKF) and is also identical to the measurement-augmentation UKF in [14]. The ECUKF is analogous to (32)−(34) for linear systems and constraints. A similar filter is explored in [5], where it is argued that the covariance of the constrained estimate is expected to be larger than that of the unconstrained estimate since the unconstrained estimate approximates the minimum variance estimate.

The two-step UKF (2UKF) [5] projects each *a posteriori* sigma point onto the constraint surface to obtain constrained sigma points. The state estimate is obtained by taking the weighted mean of the sigma points in the usual way, and the resulting estimate is then projected onto the constraint surface. Note that the mean of constrained sigma points does not itself necessarily satisfy a nonlinear constraint. 2UKF is unique in that the estimation error covariance increases after the constraints are applied. The argument for this increase is that the unconstrained estimate is the minimum variance estimate, so changing the estimate by applying constraints should increase the covariance. Furthermore, if the covariance decreases with the application of constraints (for example, using the algorithms in [13, 49]) then the covariance might become singular, which might lead to numerical problems with the matrix square root algorithm of the unscented transformation.

Unscented recursive nonlinear dynamic data reconciliation (URNDDR) [50] is similar to 2UKF. URNDDR projects the *a posteriori* sigma points onto the constraint surface, and modifies their weights based on their distances from the *a posteriori* state estimate. The modified *a posteriori* sigma points are passed through the dynamic system in the usual way to obtain the *a priori* sigma points at the next time step. The next set of *a posteriori* sigma points is obtained using a nonlinear constrained MHE with a horizon size of 1. This approach requires the solution of a nonlinear constrained optimization problem for each sigma point. The *a posteriori* state estimate and covariance are obtained by combining the sigma points in the normal way. The constraints are thus used in two different ways for the *a posteriori* estimates and covariances. URNDDR is called the sigma point interval UKF in [49]. A simplified version of URNDDR is presented in [51].

The constrained UKF (CUKF) is identical to the standard UKF, except a nonlinear constrained MHE with a horizon size of 1 is used to obtain the *a posteriori* estimate [49]. Sigma points are not projected onto the constraint surface, and constraint information is not used to modify covariances.

The constrained interval UKF (CIUKF) combines the sigma point constraints of URNDDR with the

measurement update of the CUKF [49]. That is, the CIUKF is the same as URNDDR except instead of using MHE to constrain the *a posteriori* sigma points, the unconstrained sigma points are combined to form an unconstrained estimate, and then MHE is used to constrain the estimate.

The interval UKF (IUKF) combines the post-measurement projection step of URNDDR with the measurement update of the standard unconstrained UKF [49]. That is, the IUKF is the same as URNDDR except that it skips the MHE-based constraint of the *a posteriori* sigma points. Equivalently, IUKF is also the same as CIUKF except that it skips the MHE-based constraint of the *a posteriori* state estimate.

The truncated UKF (TUKF) combines the PDF truncation approach described earlier in this paper with the UKF [49]. After each measurement update of the UKF, the PDF truncation approach is used to generated a constrained state estimate and covariance. The constrained estimate is used as the initial condition for the following time update.

The truncated interval UKF (TIUKF) adds the PDF truncation step to the *a posteriori* update of the IUKF [49]. As with the TUKF, the constrained estimate is used as the initial condition for the following time update.

In [52], the UKF is combined with MHE for constrained estimation. This is done by using a constrained UKF to estimate the first term on the right side of (76).

3.5 Interior point approaches

A new approach to inequality-constrained state estimation is called interior point likelihood maximization (IPLM) [27]. This approach is based on interior point methods, which are fundamentally different from active set methods for constraint enforcement. Active set methods for inequality constraints, as discussed earlier in this paper, proceed by solving equality-constrained subproblems and then checking if the constraints of the original problem are satisfied. One difficulty with active set methods is that computational effort grows exponentially with the number of constraints. Interior point approaches solve inequality-constrained problems by iterating using a Newton's method that is applied to a certain subproblem. The approach in [27] relies on linearization. It also has the disadvantage that the problem grows linearly with the number of time steps. However, this difficulty could possibly be addressed by limiting the horizon size, similar to MHE.

3.6 Particle filters

Particle filters operate by propagating many state estimates, called particles, that are distributed according to the PDF of the true state [2, 28]. A UKF can loosely be considered

as a type of particle filter, but UKFs and particle filters differ in several fundamental ways. First, the time update of a particle filter includes randomly generated noise that is distributed according to the known process noise PDF, while the UKF time update is deterministic. Second, the UKF has a specific number of sigma points, commonly chosen to be $n+1$ or $2n$ or $2n+1$, where n is the dimension of the state. The number of particles in a particle filter has no upper bound but typically increases exponentially with n. Third, the UKF estimates the mean and covariance of the state to third-order accuracy. The particle filter does not directly estimate the mean and covariance, but rather estimates the PDF of the state, and the PDF estimate converges to the true PDF as the number of particles approaches infinity [53]. Just as the UKF can be considered as a generalization of the EKF, the particle filter can be considered as a generalization of the UKF. Given enough particles, a particle filter always performs better than a UKF, but this might be at the expense of unacceptable computational requirements.

State-constrained particle filtering has been solved by various methods. Some of these approaches can be used with Kalman filtering, such as reparameterizing the problem [54]. Other approaches are specific to particle filtering, such as modifying the particles' likelihood functions based on their level of constraint satisfaction [55, 56] or generating process noise which ensures that the propagated particles satisfy the constraints [57]. Also, many of the methods discussed in this paper can potentially be applied to constrained particle filtering, such as projection, PDF truncation, or the SCKF. These methods could be applied to individual particles or they could be applied only to the state estimate at each time, giving rise to a large family of constrained particle filters.

Example 2: This example is taken from [14]. A discretized model of a pendulum can be written as

$$\theta_{k+1} = x_k + T\omega_k, \tag{89}$$

$$\omega_{k+1} = \omega_k - (Tg/L)\sin\theta_k, \tag{90}$$

$$y_k = \begin{bmatrix} \theta_k \\ \omega_k \end{bmatrix} + v_k, \tag{91}$$

where θ is angular position, ω is angular velocity, T is the discretization step size, g is the acceleration due to gravity, and L is the pendulum length. By conservation of energy we have

$$-mgL\cos\theta_k + mL^2\omega_k^2/2 = C, \tag{92}$$

where C is some constant. This equation is a nonlinear constraint on the states θ_k and ω_k. We use $L = 1$, trapezoidal integration with step size $T = 0.05$, $g = 9.81$, $m = 1$, and $x_0 = \begin{bmatrix} \pi/4 & \pi/50 \end{bmatrix}^T$. The covariance of the measurement noise is $R = \text{diag}(0.01, \quad 0.01)$, and the initial estimation error covariance is $P_0^+ = \text{diag}(1, \quad 1)$.

We do not use process noise in the system simulation, but in the Kalman filters we use $Q = \text{diag}(0.007^2, \quad 0.007^2)$ to help with convergence.

At this point we can take one of several approaches to the filtering problem.

1. Run the standard unconstrained EKF and ignore the constraints.

2. Linearize the constraint and run the perfect measurement EKF, or project the unconstrained EKF estimate onto the constraint surface, or use the PDF truncation method on the EKF estimate.

3. Use a second-order expansion of the constraint to project the EKF estimate onto the constraint surface.

4. Use the constrained nonlinear MHE.

5. Use the SCKF.

6. Use the UKF and either ignore the constraint, or project the *a posteriori* estimate onto the constraint surface using a linearized expansion of the constraint, or use the constrained *a posteriori* estimate to obtain the *a priori* estimate at the next time step.

7. Use the two-step UKF. Note that the corrected $\tilde{Q}$ and $\tilde{P}_0^+$, obtained using first order linearization and system projection, can also be used with these filtering approaches.

The results of MATLAB software that implements these constrained filtering algorithms [42] are shown in Table 2. The table shows the RMS state estimation errors averaged for the two states and the RMS constraint error. Each RMS value shown is averaged over 100 Monte Carlo simulations. Table 2 shows that MHE performs the best relative to estimation error. However this performance comes at a high computational expense. The Mathworks' Optimization Toolbox™ has a constrained nonlinear optimization routine called FMINCON that can be used for MHE, but for this example we use SolvOpt [58, 59]. If computational expense is a consideration then the equality constrained UKF performs the next best. However UKF implementations can also be expensive because of the sigma point calculations that are required. Table 2 shows that several of the estimators result in constraint errors that are essentially zero. The constraint errors and estimation errors are positively correlated, but small constraint errors do not guarantee that the estimation errors are small.

3.7 Summary

For nonlinear systems and nonlinear constraints, our simulation results indicate that of all the algorithms we investigated, MHE results in the smallest estimation error. However, this performance comes at the expense of programming effort and computational effort that is orders of magnitude higher than other methods. Given this caveat, it is not obvious what the "best" constrained estimation algorithm is, and it generally depends on the application. The possible approaches to constrained state estimation can be delineated by following a flowchart that

Table 2 Filter results for the nonlinear pendulum example

Filter type	RMS estimation error $(Q, \tilde{Q})$	RMS constraint error $(Q, \tilde{Q})$
unconstrained	0.0411, 0.0253	0.1167, 0.0417
perfect measurement	0.0316, 0.0905	0.0660, 0.0658
estimate projection	0.0288, 0.0207	0.0035, 0.0003
MHE, horizon size 2	0.0105, 0.0067	0.0033, 0.0008
MHE, horizon size 4	0.0089, 0.0067	0.0044, 0.0007
system projection	N/A, 0.0250	N/A, 0.0241
PDF truncation	0.0288, 0.0207	0.0035, 0.0003
2nd order constraint	0.0288, 0.0204	0.0001, 0.0000
SCKF	0.0270, 0.0235	0.0000, 0.0000
unconstrained UKF	0.0400, 0.0237	0.1147, 0.0377
projected UKF	0.0280, 0.0192	0.0046, 0.0007
equality constrained UKF	0.0261, 0.0173	0.0033, 0.0004
two-step UKF	0.0286, 0.0199	0.0005, 0.0000

Two numbers in each cell indicate the errors that are obtained when Q and $\tilde{Q}$ respectively are used in the filter. The numbers shown are RMS errors averaged over 100 Monte Carlo simulations

asks questions about the system type and the constraint type as shown in Figure 1. The acronyms used in the flowchart are given below, and the reference numbers show where the relevant equations can be found.

2E	second-order expansion of nonlinear constraints [22]
2UKF	two-step UKF [5]
CIUKF	constrained IUKF [49]
CUKF	constrained UKF [49]
ECUKF	equality constrained EKF [14]
EKF	extended Kalman filter [2]
EP	estimate projection [2]
GP	gain projection [17]
IPLM	interior point likelihood maximization [27]
IUKF	interval UKF [49]
MHE	moving horizon estimation [25, 47]
MR	model reduction [2]
PDFT	probability density function truncation [2]
PF	particle filter [2]
PM	perfect measurement [2]
PUKF	projected UKF [49]
SCKF	smoothly constrained Kalman filter [23]
SP	system projection [21]
TIUKF	truncated IUKF [49]
TUKF	truncated UKF [49]
UKF	unscented Kalman filter [2, 3]
URNDDR	unscented recursive nonlinear dynamic data reconciliation [50]

Note that some of the acronyms refer only to filter methods, some refer only to constraint incorporation
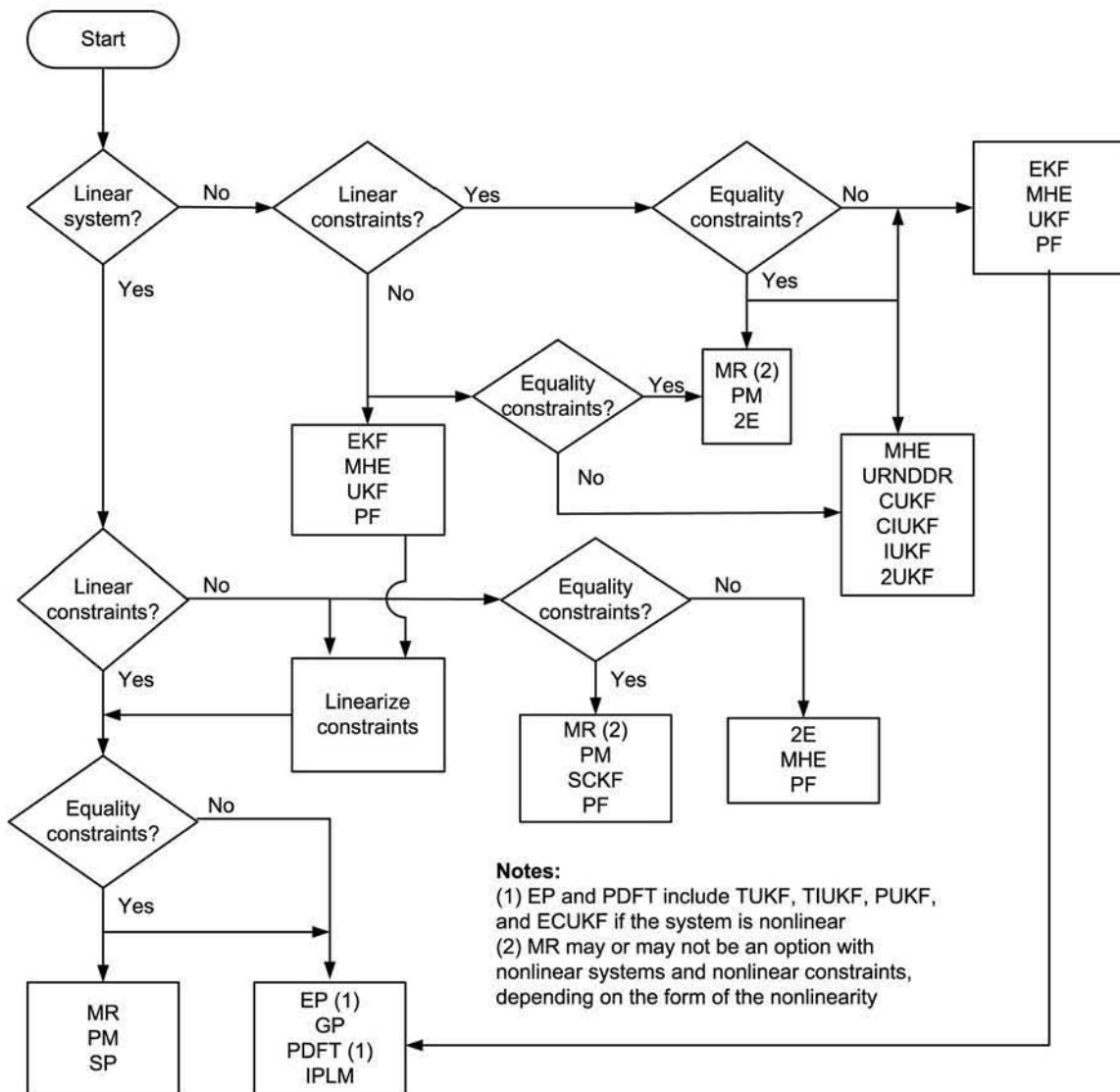

Figure 1 *Possible filter and constraint-handling choices for various combinations of system types and constraint types*

Note that some of the acronyms refer only to filter options, some refer only to constraint incorporation options, and some refer to a combination filter/constraint incorporation algorithm

methods, and some refer to a combination filter/constraint incorporation algorithm. In addition, sometimes the same acronym can refer to both a filter without constraints, and also a filter/constraint handling combination. For example, MHE can be used as an unconstrained state estimator, and then the MHE estimate can be modified by incorporating constraints using EP; or, MHE can be used as a constrained state estimator by incorporating the constraints into the MHE cost function.

4 Conclusion

The number of algorithms for constrained state estimation can be overwhelming. The reason for the proliferation of algorithms is that the problem can be viewed from many different perspectives. A linear relationship between states implies a reduction of the state dimension, hence the model reduction approach. State constraints can be viewed as perfect measurements, hence the perfect measurement approach. Constrained Kalman filtering can be viewed as a constrained likelihood maximization problem or a constrained least squares problem, hence the projection approaches. If we start with the unconstrained estimate and then incorporate the constraints to adjust the estimate we get the general projection approach and PDF truncation. If we realize that state constraints affect the relationships between the process noise terms we get the system projection approach.

Nonlinear systems and constraints have all the possibilities of nonlinear estimation, combined with all the possibilities for solving general nonlinear equations. Nonlinear systems give rise to the EKF, the UKF, MHE, and particle filtering for estimation. These estimators can be combined with various approaches for handling constraints, including first order linearization (which includes the SCKF). If first order linearization is used then any of the approaches discussed above for handling linear constraints can be used. In addition, since state estimation incorporates multiple steps (the *a priori* step and the *a posteriori* step), we can use one approach at one step and another approach at another step. The total number of possible constrained estimators seems to grow exponentially with the number of nonlinear estimation approaches and with the number of constraint handling options.

Theoretical and simulation results indicate that all of the constrained filters for linear systems and linear constraints perform identically, if the constraints are complete. Therefore in spite of the numerous approaches to the problem, we have a pleasingly parsimonious unification. However, if the constraints are not complete, then the perfect measurement and system projection methods perform best in our particular simulation example.

For nonlinear systems and nonlinear constraints, MHE resulted in the smallest estimation error in our simulation results. However, this improved performance required programming and computational effort many times higher than the other methods. The "best" constrained estimation algorithm depends on the application.

Constrained state estimation is becoming well established but we see interesting possibilities for future work. For instance, in Example 2 we see that the linearly constrained filters perform identically for the D_1 constraint, but differently for the D_2 constraint. Some of these equivalences are already proven, but conditions under which the various approaches are identical are not yet completely established. In addition, the numerical properties of the various algorithms have yet to be explored.

The second-order constraint approximation is developed in [22] and implemented in this paper in combination with the estimate projection filter. The second-order constraint approximation can also be combined with other filters, such as MHE, the UKF, and the SCKF. Algorithms for solving the second-order constraint approach can be developed and investigated for the case of multiple constraints.

More theoretical results related to convergence and stability are needed for nonlinear constrained filters such as the SCKF, MHE, and the UKF. MHE can be modified to use the optimal (singular) estimation error covariance obtained using system projection in its cost function. Second-order or iterated Kalman filtering can be combined with MHE to get a more accurate approximation of the estimation error covariance.

Various combinations of the approaches discussed in this paper can be explored. For example, PDF truncation can be combined with MHE, or the SCKF can be combined with the UKF. Conditions that are amenable to the combination of these approaches can be delineated.

The system projection approach for the nonlinear system of Example 2 uses a first-order approximation to obtain $\tilde{Q}$ and $\tilde{P}_0^+$, but a second-order approximation might give better results. In general, the extension of system projection to nonlinear systems might be a worthwhile study.

Particle filtering is a state estimation approach that is outside the scope of this paper, but it has obvious applications to constrained estimation. The theory and implementation of constrained particle filters is a topic with much room for future work.

The development of interior point methods for constrained state estimation has just begun. Further work in this area can include higher order expansion of nonlinear system and constraint equations in interior point methods, moving horizon interior point methods, the use of additional interior point theory and algorithms beyond those used in [27], and generalization of the convergence results.

It is difficult to give general guidelines for constrained filtering because each individual problem is unique. However, it appears that for either linear or nonlinear systems, the projection approaches of Sections 2.3 and 2.4 provide a good tradeoff between ease of implementation, low computational cost, and flexibility (by using different projection matrices). For soft constraints, the approaches of Section 2.8 can usually be tuned to give good performance.

For nonlinear constraints, recommended options can be listed in order of improving performance, which corresponds to greater complexity and greater computational cost. The projection approaches (Section 3) are the simplest to code and the least expensive computationally, but often provide the worst performance. These are followed in turn by the various UKF approaches (Section 3.4), MHE (Section 3.3), and finally particle filtering (Section 3.6), which can usually be expected to provide the best performance but at the highest computational cost.

The results presented in this paper can be reproduced by downloading MATLAB source code from [42].

5   References

[1]   RHODES I.: 'A tutorial introduction to estimation and filtering', *IEEE Trans. Autom. Control*, 1971, **AC-16**, (6), pp. 688−706

[2]   SIMON D.: 'Optimal state estimation' (John Wiley & Sons, 2006)

[3]   JULIER S., UHLMANN J.: 'Unscented filtering and nonlinear estimation', *Proc. IEEE*, 2004, **92**, (3), pp. 401−422

[4]   DOUCET A., DE FREITAS N., GORDON N.: 'Sequential Monte Carlo methods in practice' (Springer−Verlag, 2001)

[5]   JULIER S., LAVIOLA J.: 'On Kalman filtering with nonlinear equality constraints', *IEEE Trans. Signal Process.*, 2007, **55**, (6), pp. 2774−2784

[6]   SIMON D., SIMON D.L.: 'Kalman filtering with inequality constraints for Turbofan engine health estimation', *IEE Proc. Control Theory Appl.*, 2006, **153**, (3), pp. 371−378

[7]   VACHHANI P., RENGASWAMY R., GANGWAL V., NARASIMHAN S.: 'Recursive estimation in constrained nonlinear dynamical systems', *AIChE J.*, 2005, **51**, (3), pp. 946−959

[8]   PORRILL J.: 'Optimal combination and constraints for geometrical sensor data', *Int. J. Robot. Res.*, 1988, **7**, (6), pp. 66−77

[9]   ALOUANI A., BLAIR W.: 'Use of a kinematic constraint in tracking constant speed, maneuvering targets', *IEEE Trans. Autom. Control*, 1993, **38**, (7), pp. 1107−1111

[10]   WANG L., CHIANG Y., CHANG F.: 'Filtering method for nonlinear systems with constraints', *IEE Proc. Control Theory Appl.*, 2002, **149**, (6), pp. 525−531

[11]   CHIA T., CHOW P., CHIZEK H.: 'Recursive parameter identification of constrained systems: an application to electrically stimulated muscle', *IEEE Trans. Biomed. Eng.*, 1991, **38**, (5), pp. 429−441

[12]   SPONG M., HUTCHINSON S., VIDYASAGAR M.: 'Robot modeling and control' (John Wiley & Sons, 2005)

[13]   SIMON D., CHIA T.: 'Kalman filtering with state equality constraints', *IEEE Trans. Aerospace Electron. Syst.*, 2002, **38**, (1), pp. 128−136

[14]   TEIXEIRA B., CHANDRASEKAR J., TORRES L., AGUIRRE L., BERNSTEIN D.: 'State estimation for linear and non-linear equality-constrained systems', *Int. J. Control*, 2009, **82**, (5), pp. 918−936

[15]   WEN W., DURRANT-WHYTE H.: 'Model-based multi-sensor data fusion'. IEEE Int. Conf. on Robotics Automation, Nice, France, 1992, pp. 1720−1726

[16]   CHIA T.: 'Parameter identification and state estimation of constrained systems'. PhD thesis, Case Western Reserve University, 1985

[17]   GUPTA N., HAUSER R.: 'Kalman filtering with equality and inequality state constraints'. http://arxiv.org/abs/0709.2791, 2007, accessed May 2009

[18]   SIRCOULOMB V., ISRAEL J., HOBLOS G., CHAFOUK H., RAGOT J.: 'State estimation under nonlinear state inequality constraints. A tracking application'. 16th Mediterranean Conf. on Control Automation, Ajaccio, France, 2008, pp. 1669−1674

[19]   SHIMADA N., SHIRAI Y., KUNO Y., MIURA J.: 'Hand gesture estimation and model refinement using monocular camera − ambiguity limitation by inequality constraints'. IEEE Int. Conf. on Automatic Face Gesture Recognition, Nara, Japan, 1998, pp. 268−273

[20]   SIMON D., SIMON D.L.: 'Constrained Kalman filtering via density function truncation for turbofan engine health estimation', *Int. J. Syst. Sci.*, 2010, **41**, (2), pp. 159−171

[21]   KO S., BITMEAD R.: 'State estimation for linear systems with state equality constraints', *Automatica*, 2007, **43**, (8), pp. 1363−1368

[22]   YANG C., BLASCH E.: 'Kalman filtering with nonlinear state constraints', *IEEE Trans. Aeros. Electron. Syst.*, 2008, **45**, (1), pp. 70−84

[23]   DE GEETER J., VAN BRUSSEL H., DE SCHUTTER J.: 'A smoothly constrained Kalman filter', *IEEE Trans. Pattern Anal. Machine Intell.*, 1997, **19**, (10), pp. 1171−1177

[24] GOODWIN G., SERON M., DE DONA J.: 'Constrained control and estimation' (Springer–Verlag, 2005)

[25] RAO C., RAWLINGS J., LEE J.: 'Constrained linear state estimation – a moving horizon approach', *Automatica*, 2001, **37**, (10), pp. 1619–1628

[26] RAO C., RAWLINGS J.: 'Constrained process monitoring: moving-horizon approach', *AIChE J.*, 2002, **48**, (1), pp. 97–109

[27] BELL B., BURKE J., PILLONETTO G.: 'An inequality constrained nonlinear Kalman–Bucy smoother by interior point likelihood maximization', *Automatica*, 2009, **45**, (1), pp. 25–33

[28] ARULAMPALAM M., MASKELL S., GORDON N., CLAPP G.: 'A tutorial on particle filters for online nonlinear/non-Gaussian Bayesian tracking', *IEEE Trans. Signal Process.*, 2002, **50**, (2), pp. 174–188

[29] KALMAN R.: 'A new approach to linear filtering and prediction problems', *ASME J. Basic Eng.*, 1960, **82**, (Series D), pp. 35–45

[30] SERVI L., HO Y.: 'Recursive estimation in the presence of uniformly distributed measurement noise', *IEEE Trans. Autom. Control*, 1981, **26**, (2), pp. 563–564

[31] DORAN H.: 'Constraining Kalman filter and smoothing estimates to satisfy time-varying restrictions', *Rev. Econ. Stat.*, 1992, **74**, (3), pp. 568–572

[32] MAYBECK P.: 'Stochastic models, estimation, and control – volume 1' (Academic Press, 1979)

[33] STENGEL R.: 'Optimal control and estimation' (Dover, 1994)

[34] TUGNAIT J.: 'Constrained signal restoration via iterated extended Kalman filtering', *IEEE Trans. Acoust., Speech Signal Process.*, 1985, **ASSP-33**, (2), pp. 472–475

[35] GUPTA N.: 'Kalman filtering in the presence of state space equality constraints'. Chinese Control Conf., Harbin, China, 2007, pp. 107–113, http://arxiv.org/abs/0705.4563v1, accessed May 2009

[36] FLETCHER R.: 'Practical methods of optimization' (John Wiley & Sons, 2000)

[37] GILL P., MURRAY W., WRIGHT M.: 'Practical optimization' (Academic Press, 1981)

[38] BOYD S., VANDENBERGHE L.: 'Convex optimization' (Cambridge University Press, 2004)

[39] MAHATA K., SODERSTROM T.: 'Improved estimation performance using known linear constraints', *Automatica*, 2004, **40**, (8), pp. 1307–1318

[40] TAHK M., SPEYER J.: 'Target tracking problems subject to kinematic constraints', *IEEE Trans. Autom. Control*, 1990, **35**, (3), pp. 324–326

[41] MASSICOTTE D., MORAWSKI R., BARWICZ A.: 'Incorporation of positivity constraint into a Kalman-filter-based algorithm for correction of spectrometric data', *IEEE Trans. Instrum. Measure.*, 1995, **46**, (1), pp. 2–7

[42] SIMON, D.: 'Kalman filtering with state constraints: a survey of linear and nonlinear algorithms'. http://academic.csuohio.edu/simond/ConstrKF, accessed May 2010

[43] MICHALSKA H., MAYNE D.: 'Moving horizon observers and observer-based control', *IEEE Trans. Autom. Control*, 1995, **40**, (6), pp. 995–1006

[44] ROBERTSON D., LEE J., RAWLINGS J.: 'A moving horizon-based approach for least-squares estimation', *AIChE J.*, 1996, **42**, (8), pp. 2209–2224

[45] RUSZCZYNSKI A.: 'Nonlinear optimization' (Princeton University Press, 2006)

[46] SUN W., YUAN Y.: 'Optimization theory and methods: nonlinear programming' (Springer–Verlag, 2006)

[47] RAO C., RAWLINGS J., MAYNE D.: 'Constrained state estimation for nonlinear discrete-time systems: stability and moving horizon approximations', *IEEE Trans. Autom. Control*, 2003, **48**, (2), pp. 246–258

[48] ANDERSON T.: 'An introduction to multivariate statistical analysis' (John Wiley & Sons, 2003)

[49] TEIXEIRA B., TORRES L., AGUIRRE L., BERNSTEIN D.: 'Unscented filtering for interval-constrained nonlinear systems'. IEEE Conf. on Decision Control, Cancun, Mexico, 2008, pp. 5116–5121

[50] VACHHANI P., NARASIMHAN S., RENGASWAMY R.: 'Robust and reliable estimation via unscented recursive nonlinear dynamic data reconciliation', *J. Process Control*, 2006, **16**, (10), pp. 1075–1086

[51] KANDEPU R., IMSLAND L., FOSS B.: 'Constrained state estimation using the unscented Kalman filter'. 16th Mediterranean Conf. on Control Automation, Ajaccio, France, 2008, pp. 1453–1458

[52] QU C., HAHN J.: 'Computation of arrival cost for moving horizon estimation via unscented Kalman filtering', *J. Process Control*, 2009, **19**, (2), pp. 358–363

[53] CRISAN D., DOUCET A.: 'A survey of convergence results on particle filtering methods for practitioners', *IEEE Trans. Signal Process.*, 2002, **50**, (3), pp. 736–746

[54] AGATE C., SULLIVAN K.: 'Road-constrained target tracking and identification using a particle filter'. Signal and Data Processing of Small Targets, San Diego, CA, 2003, pp. 532–543

[55] KYRIAKIDES I., MORRELL D., PAPANDREOU-SUPPAPPOLA A.: 'A particle filtering approach to constrained motion estimation in tracking multiple targets'. Asilomar Conf. on Signals, Systems and Computers, Monterey, CA, 2005, pp. 94–98

[56] KYRIAKIDES I., MORRELL D., PAPANDREOU-SUPPAPPOLA A.: 'Multiple target tracking with constrained motion using particle filtering methods'. IEEE Int. Workshop on Computational Advances in Multi-Sensor Adaptive Processing, Puerto Vallarta, Mexico, 2005, pp. 85–88

[57] BOERS Y., DRIESSENM H.: 'Particle filter track-before-detect application using inequality constraints', *IEEE Trans. Aerosp. Electron. Syst.*, 2005, **41**, (4), pp. 1481–1487

[58] SHOR N.: 'Minimization methods for non-differentiable functions and applications' (Springer–Verlag, 1985)

[59] KUNTSEVICH, A., KAPPEL, F.: 'SolvOpt', August 1997, www.unigraz.at/imawww/kuntsevich/solvopt/, accessed May 2010