# Quadratically Constrained Least Squares with Aerospace Applications

Anton H.J. de Ruiter[1]
*Ryerson University, 350 Victoria Street, Toronto, ON, Canada M5B 2K3*

**This paper treats the problem of quadratically constrained least squares, with positive semi-definite weight matrices. A new method of solution is presented that searches directly over the constraint set, and does not require the determination of Lagrange multipliers. Global convergence of the algorithm is rigorously proven. In addition, a covariance analysis is performed for the constrained optimal solution. Two aerospace applications are presented: 1) quadratically constrained Kalman filtering similar in form to the norm-constrained Kalman filter from the literature - it is shown that the optimal quadratically constrained update is simply an orthogonal projection of the optimal unconstrained update onto the constraint set, 2) a new quadratically constrained Kalman filter using the covariance expression developed in this paper, yielding a statistically more consistent constrained filter. The new filter is demonstrated numerically with a spacecraft attitude estimation example.**

## I. Introduction

Aerospace problems such as finding an optimal sun vector from coarse sun sensor measurements [1, 2], or converting a vehicle's cartesian coordinates to geodetic coordinates [3] involve solving least squares problems with quadratic constraints. Least squares problems with quadratic constraints also arise in aerospace vehicle state estimation problems [4, 5]. Therefore, this paper treats the quadratically constrained least squares problem with positive semi-definite weight matrix.

[1] Associate Professor, Department of Aerospace Engineering; aderuiter@ryerson.ca.

To keep the presentation focussed, only literature applicable to quadratically constrained least squares will be discussed. There is significant literature on least squares problems with other types of constraints, which will not be mentioned here. A thorough study of the solutions of the quadratically constrained least squares problem has been given in [6]. In particular, it is shown that the minimizing solution corresponds to the maximum Lagrange multiplier satisfying the necessary conditions (a summary of these conditions is provided in Section II of this paper). It is further shown that the Lagrange multipliers satisfy a rational equation, which must then be solved for the maximum root. The conventional method for solving the quadratic least squares problem is based on this, by using a Newton method to solve the rational equation for the Lagrange multipliers [7, 8]. As discussed in [9], the rational equation can become quite difficult to solve, since the maximum root can be quite close to a pole of the same equation. As such, convergence of Newton's method for solving it can be slow. In [10, 11], procedures were presented for approximately determining the maximum root when the weight matrix is strictly positive-definite, under the assumption that the maximum Lagrange multiplier is positive. However, as demonstrated in Section II, this places a restriction on the quadratically constrained least squares problems that can be solved in this way. Recently, it was demonstrated in [5] that the finding the roots of the rational equation is equivalent to finding the eigenvalues of an associated companion matrix. This is accomplished by writing the rational function in terms of a common denominator, and then setting the numerator, which is a polynomial, equal to zero. The companion matrix is formed by computing the coefficients for the polynomial and using those to construct a matrix with the same characteristic equation. However, this still requires the construction of the companion matrix, and then finding its eigenvalues.

The objective in this paper is to develop a method for solving the quadratically constrained least squares problem that is suitable for online implementation on a vehicle with possibly limited computing capabilities. The conventional solution procedure based on computing the maximum Lagrange multiplier could become computationally cumbersome particularly for high dimensional problems, and may not be suitable for implementation on a vehicle's processor in real-time, even in the case of low dimensional problems if a vehicle's processor has limited computational capacity such as is often the case on very small satellites. To overcome the need for computing the Lagrange multiplier entirely, [9] proposed a method of optimization by searching directly over the constraint set. Reference [9] restricted the attention to least squares with a unit sphere constraint (a special case of the problem considered in this paper). In particular, the method in [9] projects the steepest descent direction of the cost function onto the tangent space of the sphere, and then performs a line search along the corresponding geodesic. For a unit sphere, the geodesic has a very simple form, which is not the case for the more general constraint set considered in this paper. Furthermore, the steepest descent direction may not provide the best rate of convergence. For instance, in unconstrained minimization problems, it is known that steepest descent algorithms provide better initial convergence for large deviations from the minimum, while Newton-type algorithms provide better convergence in the neighborhood of the minimum [12]. As such, this paper generalizes the idea from [9] to a quadratic constraint set with positive semi-definite weight matrix, with a search curve defined as a convex combination between steepest descent and Newton-type directions. Unlike [9], the search curve does not require the computation of a geodesic on the constraint set.

The remainder of this paper is organized as follows. Section II contains the problem formulation together with a further discussion of the shortcomings of existing solution methods. Section III contains the main contribution of this paper, which is the new algorithm for solving the quadratically constrained least squares problem directly, without the need for computing Lagrange multipliers or making any coordinate transformations. Section III also contains a covariance analysis of the constrained solution, which is another contribution of this paper. Section IV presents two applications of the proposed method. The first application in Section IV.A is quadratically constrained Kalman filtering of the form in [4, 5]. It is shown here that for such a filter with a quadratic state estimate constraint, the constrained Kalman filter reduces to an orthogonal projection of the unconstrained Kalman estimate onto the constraint surface, analogous to the norm-constrained case in [4] - another contribution of this paper. This orthogonal projection is readily performed using the method in this paper. The second application in Section IV.B is an alternative statistically more consistent formulation of a quadratically constrained Kalman filter - another contribution of this paper. Specifically, for the constrained Kalman filters as in Section IV.A of this paper and [4, 5], the covariance-like matrix $\mathbf{P}$ being propagated loses the meaning of the state estimate covariance, since it neglects the fact that the constrained Kalman gain is dependent on the measurement residual [4]. In Section IV.B, a different formulation of a quadratically constrained Kalman filter is presented, using the covariance expression developed in Section III of this paper. In this way, the matrix $\mathbf{P}$ being propagated maintains the interpretation of the state estimate covariance. Finally, Section V contains the conclusions, while the three appendices contain the detailed proofs of the mathematical results presented this paper.

## II. Problem Formulation

The following notation will be used throughout the paper. The $n \times n$ identity matrix will be denoted by $\mathbf{I}_n$, while $\mathbf{0}_{n\times m}$ denotes the $n \times m$ matrix of zeros. However, the dimension will not be indicated if it is clear from context. For $\mathbf{x} \in \mathbb{R}^n$, $\|\mathbf{x}\|_2$ denotes the Euclidean norm of $\mathbf{x}$, while $\|\mathbf{X}\|_2$ denotes the corresponding induced norm for the matrix $\mathbf{X} \in \mathbb{R}^{m\times n}$, which is given by $\|\mathbf{X}\|_2 = \bar{\sigma}(\mathbf{X})$, where $\bar{\sigma}(\cdot)$ denotes the largest singular value.

This paper treats the following quadratically constrained least squares problem:

**Problem 1**

Find $\mathbf{x} \in \mathbb{R}^n$ to minimize

$$J(\mathbf{x}) = \frac{1}{2}\left(\mathbf{y} - \mathbf{H}\mathbf{x}\right)^T \mathbf{W}\left(\mathbf{y} - \mathbf{H}\mathbf{x}\right) \tag{1}$$

subject to

$$\mathbf{x}^T\mathbf{D}\mathbf{x} = \ell, \tag{2}$$

where $\mathbf{y} \in \mathbb{R}^m$ with $m \geq n$, $\mathbf{W} = \mathbf{W}^T > \mathbf{0}$, $\mathbf{D} = \mathbf{D}^T \geq \mathbf{0}$, $\ell > 0$, and $\mathbf{H} \in \mathbb{R}^{m\times n}$ has full column rank.

It can be shown that the unconstrained minimizing solution of (1) is given by [12]

$$\mathbf{x}_{unc} = \overline{\mathbf{W}}^{-1}\mathbf{H}^T\mathbf{W}\mathbf{y}, \;\; \overline{\mathbf{W}} = \mathbf{H}^T\mathbf{W}\mathbf{H}, \tag{3}$$

where $\overline{\mathbf{W}}$ is also positive definite. Given the unconstrained minimizer of $J(\mathbf{x})$ in (3), it can be shown that the original minimization problem (Problem 1) is equivalent to:

**Problem 2**

Minimize

$$\bar{J}(\mathbf{x}) = \frac{1}{2}\left(\mathbf{x}_{unc} - \mathbf{x}\right)^T \overline{\mathbf{W}}\left(\mathbf{x}_{unc} - \mathbf{x}\right). \tag{4}$$

subject to

$$\mathbf{x}^T\mathbf{D}\mathbf{x} = \ell, \tag{5}$$

where $\mathbf{x}_{unc}$ and $\overline{\mathbf{W}}$ are given in (3).

The conventional Lagrange multiplier method for solving Problem 1 is now outlined. The Lagrangian for Problem 2 is given by [13]

$$L(\mathbf{x},\lambda) = \frac{1}{2}\left(\mathbf{x}_{unc} - \mathbf{x}\right)^T \overline{\mathbf{W}}\left(\mathbf{x}_{unc} - \mathbf{x}\right) + \frac{\lambda}{2}\left(\mathbf{x}^T\mathbf{D}\mathbf{x} - \ell\right), \tag{6}$$

where $\lambda$ is the Lagrange multiplier. Since the constraint gradient, $\mathbf{D}\mathbf{x}$, is non-zero at the minimum by (5), the necessary condition for a minimum is [13]

$$\frac{\partial L}{\partial \mathbf{x}} = -\overline{\mathbf{W}}\left(\mathbf{x}_{unc} - \mathbf{x}\right) + \lambda\mathbf{D}\mathbf{x} = \mathbf{0}, \tag{7}$$

together with (5). Solving (7), gives

$$\mathbf{x} = \left(\lambda\mathbf{D} + \overline{\mathbf{W}}\right)^{-1}\overline{\mathbf{W}}\mathbf{x}_{unc}. \tag{8}$$

Substituting this into the constraint equation (5), leads to

$$\mathbf{x}^T\mathbf{D}\mathbf{x} = \mathbf{x}_{unc}^T\overline{\mathbf{W}}\left(\lambda\mathbf{D} + \overline{\mathbf{W}}\right)^{-1}\mathbf{D}\left(\lambda\mathbf{D} + \overline{\mathbf{W}}\right)^{-1}\overline{\mathbf{W}}\mathbf{x}_{unc} = \ell. \tag{9}$$

Since $\overline{\mathbf{W}}$ is symmetric and positive definite, and $\mathbf{D}$ is symmetric and positive semi-definite, there exists a matrix $\mathbf{E} \in \mathbb{R}^{n\times n}$ such that [14]

$$\mathbf{E}^T\overline{\mathbf{W}}\mathbf{E} = \mathbf{I},\ \ \mathbf{E}^T\mathbf{D}\mathbf{E} = \mathbf{\Sigma}, \tag{10}$$

where $\mathbf{\Sigma} = \operatorname{diag}\{\sigma_1, ..., \sigma_n\}$, and $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_n \geq 0$. Making use of this simultaneous diagonalization, the constraint equation (9) becomes

$$\mathbf{x}_{unc}^T\mathbf{F}^T\mathbf{M}\mathbf{F}\mathbf{x}_{unc} = \ell, \tag{11}$$

where $\mathbf{F} = \mathbf{E}^{-1}$, and $\mathbf{M} = \operatorname{diag}\left\{\sigma_1/(1+\sigma_1\lambda)^2, ..., \sigma_n/(1+\sigma_n\lambda)^2\right\}$. Setting $\begin{bmatrix} z_1 & \cdots & z_n \end{bmatrix}^T = \mathbf{F}\mathbf{x}_{unc}$, the constraint equation (11) becomes [6]

$$\sum_{i=1}^{n} \frac{z_i^2\sigma_i}{(1+\sigma_i\lambda)^2} = \ell. \tag{12}$$

Equation (12) must now be solved numerically to find the maximum root $\lambda$, from which the constrained minimizing solution in (8) may be computed [6]. Further complicating matters, the development of (12) assumes that for the Lagrange multipliers $\lambda$, the matrix $\lambda\mathbf{D} + \overline{\mathbf{W}}$ is invertible. This condition is violated when $\overline{\mathbf{W}}\mathbf{x}_{unc} = (-\mu\mathbf{D} + \overline{\mathbf{W}})\bar{\mathbf{x}}$ for some $\bar{\mathbf{x}} \in \mathbb{R}^n$, such that $\rho^2\mathbf{e}^T\mathbf{D}\mathbf{e} + 2\rho\bar{\mathbf{x}}^T\mathbf{D}\mathbf{e} + \bar{\mathbf{x}}^T\mathbf{D}\bar{\mathbf{x}} = \ell$, has a real solution $\rho \in \mathbb{R}$ where $\mu$ is a generalized eigenvalue of $\mathbf{D}, \overline{\mathbf{W}}$, with corresponding generalized eigenvector $\mathbf{e} \in \mathbb{R}^n$. In this case, (7) is satisfied with $\lambda = -\mu$ and $\mathbf{x} = \bar{\mathbf{x}} + \rho\mathbf{e}$. Further analysis of this scenario using the simultaneous diagonalization in (10) shows that a generalized eigenvalue $\mu$ of $\mathbf{D}, \overline{\mathbf{W}}$ yields a valid Lagrange multiplier $\lambda = -\mu$ if and only if $\mathbf{x}_{unc}$ is small enough (specifically $\bar{\mathbf{x}}^T\mathbf{D}\bar{\mathbf{x}} \leq \ell$) and lies in the subspace spanned by the generalized eigenvectors of $\mathbf{D}, \overline{\mathbf{W}}$ corresponding to the generalized eigenvalues distinct from $\mu$. Therefore, this situation also needs to be checked to ensure that no Lagrange multipliers are missed. Even if $\lambda\mathbf{D} + \overline{\mathbf{W}}$ is invertible for the maximum Lagrange multiplier, the minimizing solution $\mathbf{x}$ in (8) can be very sensitive to errors in $\lambda$ when it is close to a generalized eigenvalue of $\mathbf{D}, \overline{\mathbf{W}}$.

In [10, 11], procedures are developed for approximately determining the maximum root of (12) when $\mathbf{D}$ is strictly positive-definite, under the assumption that this maximum root is positive. However, as can be shown, if $\mathbf{x}_{unc}^T\mathbf{D}\mathbf{x}_{unc} \leq \ell$, no positive solution $\lambda$ exists. Indeed, premultiplying (7) by $\mathbf{x}^T\mathbf{D}\overline{\mathbf{W}}^{-1}$, and rearranging gives

$$\lambda = \mathbf{x}^T\mathbf{D}(\mathbf{x}_{unc} - \mathbf{x}) / \left(\mathbf{x}^T\mathbf{D}\overline{\mathbf{W}}^{-1}\mathbf{D}\mathbf{x}\right). \tag{13}$$

Note that the denominator is positive by the constraint $\mathbf{x}^T\mathbf{D}\mathbf{x} = \ell$, and the positive-definiteness of $\overline{\mathbf{W}}^{-1}$. Now, suppose that $\mathbf{x}_{unc}^T\mathbf{D}\mathbf{x}_{unc} \leq \ell$. Then,

$$\begin{aligned} 0 &\leq (\mathbf{x}_{unc} - \mathbf{x})^T\mathbf{D}(\mathbf{x}_{unc} - \mathbf{x}) = \mathbf{x}_{unc}^T\mathbf{D}\mathbf{x}_{unc} + \mathbf{x}^T\mathbf{D}\mathbf{x} - 2\mathbf{x}^T\mathbf{D}\mathbf{x}_{unc}, \\ &\leq 2(\ell - \mathbf{x}^T\mathbf{D}\mathbf{x}_{unc}) = -2\mathbf{x}^T\mathbf{D}(\mathbf{x}_{unc} - \mathbf{x}). \end{aligned}$$

Substituting this into (13) shows that necessarily $\lambda \leq 0$ in this case. Suppose that, as in many practical aerospace problems, $\mathbf{y}$ in Problem 1 is a measurement that satisfies $\mathbf{y} = \mathbf{H}\mathbf{x}_{true} + \mathbf{v}$, where $\mathbf{x}_{true}$ is the true value of $\mathbf{x}$, which satisfies $\mathbf{x}_{true}^T\mathbf{D}\mathbf{x}_{true} = \ell$, and $\mathbf{v}$ is a random measurement error. In general, for such $\mathbf{y}$ there is no guarantee that the unconstrained solution $\mathbf{x}_{unc}$ in (3) satisfies $\mathbf{x}_{unc}^T\mathbf{D}\mathbf{x}_{unc} > \ell$. Therefore the the maximum root $\lambda$ of (12) may not be positive, and the solution procedures in [10, 11] would not apply.

## III. Main Results

To develop the solution procedure, let a trial point $\mathbf{x}_k$ satisfying $\mathbf{x}_k^T \mathbf{D} \mathbf{x}_k = \ell$ be given for some $k \geq 0$. Assuming that $\mathbf{x}_k$ is not a solution of Problem 1, the objective is to find a new point $\mathbf{x}_{k+1}$ satisfying $\mathbf{x}_{k+1}^T \mathbf{D} \mathbf{x}_{k+1} = \ell$, with $\bar{J}(\mathbf{x}_{k+1}) < \bar{J}(\mathbf{x}_k)$. To this end, consider the search curve

$$\mathbf{x}_k^s(\alpha) = (\mathbf{x}_k + \alpha \mathbf{b}_k) \sqrt{\frac{\ell}{\ell + \alpha^2 b}}, \tag{14}$$

for $\alpha \geq 0$, where

$$\mathbf{b}_k = \left[ \beta_k \left( \mathbf{I} - \frac{\mathbf{D}\mathbf{x}_k \mathbf{x}_k^T \mathbf{D}}{\|\mathbf{D}\mathbf{x}_k\|_2^2} \right) \overline{\mathbf{W}} + (1 - \beta_k) \left( \mathbf{I} - \frac{\overline{\mathbf{W}}^{-1} \mathbf{D}\mathbf{x}_k \mathbf{x}_k^T \mathbf{D}}{\mathbf{x}_k^T \mathbf{D} \overline{\mathbf{W}}^{-1} \mathbf{D} \mathbf{x}_k} \right) \right] (\mathbf{x}_{unc} - \mathbf{x}_k), \tag{15}$$

$$b = \mathbf{b}_k^T \mathbf{D} \mathbf{b}_k, \tag{16}$$

and $0 \leq \beta_k \leq 1$ is some user defined weighting parameter. It can be shown that $\mathbf{x}_k^s(\alpha)^T \mathbf{D} \mathbf{x}_k^s(\alpha) = \ell$ for all $\alpha \geq 0$. As such, the search curve lies in the constraint set.

It can be shown that $\mathbf{b}_k^T \mathbf{D} \mathbf{x}_k = \mathbf{0}$. As such, $\mathbf{b}_k$ lies in the tangent space of the constraint set at the point $\mathbf{x}_k$. When $\beta_k = 1$, it is seen that $\mathbf{b}_k$ is a projection of the negative gradient of the cost function in (4) onto the tangent space of the constraint set. As such, with $\beta_k = 1$, the search along the curve (14) can be thought of as a steepest descent search on the constraint set. When $\beta_k = 0$, the search curve can be given the following interpretation. First, for a general unconstrained optimization problem with cost $J(\mathbf{x})$, a Newton search direction is given by $-(\nabla^2 J)^{-1} \nabla J$. For the cost function in (4), $\nabla^2 \bar{J} = \overline{\mathbf{W}}$ and the Newton search direction would be $\mathbf{x}_{unc} - \mathbf{x}_k$. However, the projection of the Newton search direction onto the tangent space of the constraint set is not necessarily a descent direction. Therefore, the projection must be modified, as shown in the second term of (15). As such, with $\beta_k = 0$, the search along the curve (14) can be thought of as a Newton type search on the constraint set. Finally, it is known that steepest descent algorithms provide better initial convergence for large deviations from the minimum, while Newton type algorithms provide better convergence in the neighborhood of the minimum [12]. Similar to the Levenberg-Marquardt method [12], the parameter $\beta_k$ can be appropriately chosen by the user to improve the convergence of the optimization algorithm for large initial deviations from the minimum.

The cost function (4) along $\mathbf{x}_k^s(\alpha)$ is

$$\bar{J}(\mathbf{x}_k^s(\alpha)) = \frac{1}{2}\left(\mathbf{x}_{unc} - \mathbf{x}_k^s(\alpha)\right)^T \overline{\mathbf{W}}\left(\mathbf{x}_{unc} - \mathbf{x}_k^s(\alpha)\right), \tag{17}$$

A new function yielding the value of the cost function along the search curve $\mathbf{x}_k^s(\alpha)$ is defined as

$$\phi_k(\alpha) = \bar{J}(\mathbf{x}_k^s(\alpha)). \tag{18}$$

Note that $\mathbf{x}_k^s(0) = \mathbf{x}_k$, and correspondingly,

$$\bar{J}(\mathbf{x}_k) = \phi_k(0). \tag{19}$$

Differentiating (14), yields

$$\frac{d\mathbf{x}_k^s}{d\alpha} = \mathbf{b}_k\sqrt{\frac{\ell}{\ell+\alpha^2 b}} - \frac{\alpha b}{l+\alpha^2 b}\mathbf{x}_k^s(\alpha). \tag{20}$$

As a result, differentiating (18) leads to

$$\frac{d\phi_k}{d\alpha} = -\left(\mathbf{x}_{unc} - \mathbf{x}_k^s(\alpha)\right)^T \overline{\mathbf{W}}\mathbf{b}_k\sqrt{\frac{\ell}{\ell+\alpha^2 b}} + \frac{\alpha b}{\ell+\alpha^2 b}\left(\mathbf{x}_{unc} - \mathbf{x}_k^s(\alpha)\right)^T \overline{\mathbf{W}}\mathbf{x}_k^s(\alpha). \tag{21}$$

In particular, for $\alpha = 0$, using (15) one obtains

$$\begin{aligned}\frac{d\phi_k}{d\alpha}(0) \;&=\; -\beta_k\left(\mathbf{x}_{unc} - \mathbf{x}_k\right)^T \overline{\mathbf{W}}\left(\mathbf{I} - \frac{\mathbf{D}\mathbf{x}_k\mathbf{x}_k^T\mathbf{D}}{\|\mathbf{D}\mathbf{x}_k\|_2^2}\right)\overline{\mathbf{W}}\left(\mathbf{x}_{unc} - \mathbf{x}_k\right)\\ &\quad -(1-\beta_k)\left(\mathbf{x}_{unc} - \mathbf{x}_k\right)^T\left(\overline{\mathbf{W}} - \frac{\mathbf{D}\mathbf{x}_k\mathbf{x}_k^T\mathbf{D}}{\mathbf{x}_k^T\mathbf{D}\overline{\mathbf{W}}^{-1}\mathbf{D}\mathbf{x}_k}\right)\left(\mathbf{x}_{unc} - \mathbf{x}_k\right).\end{aligned} \tag{22}$$

Using the simultaneous diagonalization in (10), the central matrix in the second line of (22) can be written as

$$\overline{\mathbf{W}} - \frac{\mathbf{D}\mathbf{x}_k\mathbf{x}_k^T\mathbf{D}}{\mathbf{x}_k^T\mathbf{D}\overline{\mathbf{W}}^{-1}\mathbf{D}\mathbf{x}_k} = \mathbf{E}^{-T}\left(\mathbf{I} - \hat{\mathbf{n}}\hat{\mathbf{n}}^T\right)\mathbf{E}^{-1}, \tag{23}$$

where $\hat{\mathbf{n}} = \boldsymbol{\Sigma}\mathbf{E}^{-1}\mathbf{x}_k/\|\boldsymbol{\Sigma}\mathbf{E}^{-1}\mathbf{x}_k\|$. Consequently, this matrix is positive-semidefinite, and from (22)

$$\frac{d\phi_k}{d\alpha}(0) \leq 0. \tag{24}$$

**Proposition 1**

Let $\mathbf{x}_k$ satisfy $\mathbf{x}_k^T\mathbf{D}\mathbf{x}_k = \ell$. Then, equality in (24) holds if and only if $\mathbf{x}_k$ satisfies the necessary condition in (7).

**Proof** See Appendix A 1.

Proposition 1 together with (24) show that if $\mathbf{x}_k$ is not a stationary point of Problem 1, then $d\phi_k/d\alpha(0) < 0$, and $\phi_k(\alpha)$ is strictly decreasing at $\alpha = 0$. The following useful preliminary result is now obtained:

**Proposition 2**

Let $\phi_k(\alpha)$ be defined as in (18), and consider the level set $\mathcal{L} = \{\mathbf{x} \in \mathbb{R}^n : \bar{J}(\mathbf{x}) \leq \bar{J}_0\}$, for some $\bar{J}_0 > \phi_k(0)$. Then, there exists an $M > 0$, such that for all $\mathbf{x}_k \in \text{int}[\mathcal{L}]$ satisfying $\mathbf{x}_k^T \mathbf{D} \mathbf{x}_k = \ell$,

$$\left| \frac{d\phi_k}{d\alpha}(\alpha_1) - \frac{d\phi_k}{d\alpha}(\alpha_2) \right| \leq M \left| \alpha_1 - \alpha_2 \right|, \quad \alpha_1, \alpha_2 \in [0, \bar{\alpha}). \tag{25}$$

where $\bar{\alpha} > 0$ is either the smallest value of $\alpha$ such that $\phi_k(\alpha) = \bar{J}_0$ if it exists, or $\bar{\alpha} = +\infty$, and $\text{int}[\mathcal{L}] = \{\mathbf{x} \in \mathbb{R}^n : \bar{J}(\mathbf{x}) < \bar{J}_0\}$ (the interior of $\mathcal{L}$).

**Proof** See Appendix A 2.

Given a non-stationary trial point $\mathbf{x}_k$ and a step length $\alpha_k$ for $k \geq 0$, set

$$\mathbf{x}_{k+1} = \mathbf{x}_k^s(\alpha_k), \tag{26}$$

for some appropriately chosen step size $\alpha_k > 0$. Proposition 2 is a key property that allows one to require the chosen step size to satisfy the Wolfe conditions [13]. Namely, given constants $c_1$ and $c_2$ satisfying $0 < c_1 < c_2 < 1$, the chosen step size is required to satisfy

$$\phi_k(\alpha_k) \leq \phi_k(0) + \alpha_k c_1 \frac{d\phi_k}{d\alpha}(0) \tag{27}$$

$$\frac{d\phi_k}{d\alpha}(\alpha_k) \geq c_2 \frac{d\phi_k}{d\alpha}(0) \tag{28}$$

It can be shown (see for example [13]) that such a step length exists. Furthermore, there are algorithms (see for example [13]), which find such step lengths. The following algorithm is now obtained for solving Problem 1:

**Algorithm 1**

Compute $\mathbf{x}_{unc}$ from (3). Let $\alpha_{\max} > 0$ be a user defined maximum step length, and let $\mathbf{x}_0 \in \mathbb{R}^n$ satisfy $\mathbf{x}_0^T \mathbf{D} \mathbf{x}_0 = \ell$. For $k \geq 0$:

1. Compute $d\phi_k/d\alpha(0)$ using (22). If $d\phi_k/d\alpha(0) = 0$, then set $\mathbf{x}_{k+1} = \mathbf{x}_k$. Else, find a step length $\alpha_k \in (0, \alpha_{\max}]$ satisfying (27) and (28). If no such step length exists, set $\alpha_k = \alpha_{\max}$. Compute $\mathbf{x}_{k+1} = \mathbf{x}_k^s(\alpha_k)$.

2. Set $k \to k+1$ and return to 1.

The following convergence result is now obtained:

**Theorem 1** Consider Algorithm 1, and define the sets $\mathcal{J} = \{\mathbf{x} \in \mathbb{R}^n : \mathbf{x}^T \mathbf{D} \mathbf{x} = \ell, J(\mathbf{x}) \leq J(\mathbf{x}_0)\}$, and $\mathcal{C}$ the set of stationary points of Problem 1. Then, $\mathbf{x}_k \in \mathbb{R}^n$ approaches the set $\mathcal{J} \bigcap \mathcal{C}$. Furthermore, $\mathbf{x}_{k+1} - \mathbf{x}_k \to \mathbf{0}$ as $k \to \infty$.

**Proof** See Appendix A 3.

A reasonable choice for $\mathbf{x}_0$ is $\mathbf{x}_0 = \mathbf{x}_{unc}\sqrt{\ell/\mathbf{x}_{unc}^T \mathbf{D} \mathbf{x}_{unc}}$. This is a particularly good choice if one expects that the unconstrained solution $\mathbf{x}_{unc}$ is close to the constraint surface, such as would be the case if $\mathbf{y}$ in Problem 1 are measurements generated by $\mathbf{y} = \mathbf{H}\mathbf{x}_{true} + \mathbf{v}$, where $\mathbf{x}_{true}$ satisfies $\mathbf{x}_{true}^T \mathbf{D} \mathbf{x}_{true} = \ell$ and $\mathbf{v}$ is some small unknown measurement error.

### A. Sensitivity Analysis

Suppose that the measurement satisfies

$$\mathbf{y} = \mathbf{H}\mathbf{x}_{true} + \mathbf{v}, \tag{29}$$

where $\mathbf{x}_{true}$ is the true value of $\mathbf{x}$ satisfying $\mathbf{x}_{true}^T \mathbf{D} \mathbf{x}_{true} = \ell$, and $\mathbf{v}$ is some small unknown measurement error. This section examines the sensitivity of the solution $\mathbf{x}$ of Problem 1 to small variations in $\mathbf{v}$.

From the necessary conditions in (7),

$$\lambda \mathbf{D}\mathbf{x} + \overline{\mathbf{W}}\mathbf{x} = \overline{\mathbf{W}}\mathbf{x}_{unc}. \tag{30}$$

Substituting the unconstrained least squares solution $\mathbf{x}_{unc}$ from (3) into (30) and making use of

(29), leads to

$$\lambda \mathbf{D}\mathbf{x} + \overline{\mathbf{W}}\mathbf{x} = \overline{\mathbf{W}}\mathbf{x}_{true} + \mathbf{H}^T\mathbf{W}\mathbf{v}. \tag{31}$$

Pre-multiplying both sides of (31) by $\mathbf{x}^T$ and making use of the constraint $\mathbf{x}^T\mathbf{D}\mathbf{x} = \ell$, yields the Lagrange multiplier

$$\lambda = \frac{1}{\ell}\mathbf{x}^T\overline{\mathbf{W}}\left(\mathbf{x}_{true} - \mathbf{x}\right) + \frac{1}{\ell}\mathbf{x}^T\mathbf{H}^T\mathbf{W}\mathbf{v}. \tag{32}$$

Defining the estimation error as

$$\delta\mathbf{x} = \mathbf{x} - \mathbf{x}_{true}, \tag{33}$$

the Lagrange multiplier becomes

$$\lambda = \frac{1}{\ell}\left(-\mathbf{x}_{true}^T\overline{\mathbf{W}}\delta\mathbf{x} - \delta\mathbf{x}^T\overline{\mathbf{W}}\delta\mathbf{x} + \mathbf{x}_{true}^T\mathbf{H}^T\mathbf{W}\mathbf{v} + \delta\mathbf{x}^T\mathbf{H}^T\mathbf{W}\mathbf{v}\right). \tag{34}$$

Likewise, substituting (33) into (31) gives

$$\lambda\mathbf{D}\mathbf{x}_{true} + \left(\lambda\mathbf{D} + \overline{\mathbf{W}}\right)\delta\mathbf{x} = \mathbf{H}^T\mathbf{W}\mathbf{v} \tag{35}$$

Substituting (34) into (35) and retaining only terms up to first order in $\delta\mathbf{x}$ and $\mathbf{v}$, gives

$$\left(\mathbf{I} - \frac{1}{\ell}\mathbf{D}\mathbf{x}_{true}\mathbf{x}_{true}^T\right)\overline{\mathbf{W}}\delta\mathbf{x} = \left(\mathbf{I} - \frac{1}{\ell}\mathbf{D}\mathbf{x}_{true}\mathbf{x}_{true}^T\right)\mathbf{H}^T\mathbf{W}\mathbf{v}. \tag{36}$$

From the constraint $\mathbf{x}^T\mathbf{D}\mathbf{x} = \ell$, one has to first order,

$$\mathbf{x}_{true}^T\mathbf{D}\delta\mathbf{x} = 0. \tag{37}$$

Combining this with (36), leads to

$$\mathbf{A}\delta\mathbf{x} = \begin{bmatrix} \left(\mathbf{I} - \frac{1}{\ell}\mathbf{D}\mathbf{x}_{true}\mathbf{x}_{true}^T\right)\mathbf{H}^T\mathbf{W}\mathbf{v} \\ 0 \end{bmatrix}, \tag{38}$$

where

$$\mathbf{A} = \begin{bmatrix} \left(\mathbf{I} - \frac{1}{\ell}\mathbf{D}\mathbf{x}_{true}\mathbf{x}_{true}^T\right)\overline{\mathbf{W}} \\ \zeta\mathbf{x}_{true}^T\mathbf{D} \end{bmatrix}, \tag{39}$$

with $\zeta \in \mathbb{R}$ and $\zeta \neq 0$.

It is now shown that $\mathbf{A}$ in (39) has full column rank. First, note that $\mathbf{x}_{true}$ and $\mathbf{D}\mathbf{x}_{true}$ are non-zero, due to the constraint $\mathbf{x}_{true}^T\mathbf{D}\mathbf{x}_{true} = \ell$. Next, let $\mathbf{z}_1, ..., \mathbf{z}_{n-1} \in \mathbb{R}^n$ extend $\mathbf{x}_{true}$ to an orthogonal basis for $\mathbb{R}^n$. In a similar manner to the proof of Proposition 1, it can be shown that $\mathbf{D}\mathbf{x}_{true}, \mathbf{z}_1, ..., \mathbf{z}_{n-1}$ also forms a basis for $\mathbb{R}^n$. Now, suppose that

$$\mathbf{A}\mathbf{q} = \mathbf{0}, \tag{40}$$

for some $\mathbf{q} \in \mathbb{R}^n$. One can write

$$\overline{\mathbf{W}}\mathbf{q} = \gamma\mathbf{D}\mathbf{x}_{true} + \sum_{i=1}^{n-1} \alpha_i \mathbf{z}_i,$$

for some $\gamma, \alpha_1, ..., \alpha_{n-1} \in \mathbb{R}$. From (39) and (40), one obtains

$$\left(\mathbf{I} - \frac{1}{\ell}\mathbf{D}\mathbf{x}_{true}\mathbf{x}_{true}^T\right)\left(\gamma\mathbf{D}\mathbf{x}_{true} + \sum_{i=1}^{n-1} \alpha_i \mathbf{z}_i\right) = \mathbf{0}, \tag{41}$$

$$\mathbf{x}_{true}^T\mathbf{D}\mathbf{q} = 0. \tag{42}$$

By the orthogonality of $\mathbf{x}_{true}$ and $\mathbf{z}_1, ..., \mathbf{z}_{n-1}$, and the condition $\mathbf{x}_{true}^T\mathbf{D}\mathbf{x}_{true} = \ell$, equation (41) leads to $\sum_{i=1}^{n-1} \alpha_i \mathbf{z}_i = 0$, and hence $\overline{\mathbf{W}}\mathbf{q} = \gamma\mathbf{D}\mathbf{x}_{true}$. Substituting this into (42) gives $\gamma\mathbf{x}_{true}^T\mathbf{D}\overline{\mathbf{W}}^{-1}\mathbf{D}\mathbf{x}_{true} = 0$, leading to $\gamma = 0$. Therefore, the only solution to (40) is the trivial solution $\mathbf{q} = \mathbf{0}$, which implies that $\mathbf{A}$ has full column rank.

Consequently, (38) can be solved as

$$\delta\mathbf{x} = \left(\mathbf{A}^T\mathbf{A}\right)^{-1}\mathbf{A}^T\begin{bmatrix}\left(\mathbf{I} - \frac{1}{\ell}\mathbf{D}\mathbf{x}_{true}\mathbf{x}_{true}^T\right)\mathbf{H}^T\mathbf{W}\mathbf{v} \\ 0\end{bmatrix}. \tag{43}$$

Evaluating (43) gives

$$\delta\mathbf{x} = \mathbf{B}\mathbf{H}^T\mathbf{W}\mathbf{v}, \tag{44}$$

where

$$\begin{aligned}\mathbf{B} = &\left(\overline{\mathbf{W}}\left(\mathbf{I} - \frac{1}{\ell}\mathbf{D}\mathbf{x}_{true}\mathbf{x}_{true}^T\right)^T\left(\mathbf{I} - \frac{1}{\ell}\mathbf{D}\mathbf{x}_{true}\mathbf{x}_{true}^T\right)\overline{\mathbf{W}} + \zeta^2\mathbf{D}\mathbf{x}_{true}\mathbf{x}_{true}^T\mathbf{D}\right)^{-1} \\ &\times\overline{\mathbf{W}}\left(\mathbf{I} - \frac{1}{\ell}\mathbf{D}\mathbf{x}_{true}\mathbf{x}_{true}^T\right)^T\left(\mathbf{I} - \frac{1}{\ell}\mathbf{D}\mathbf{x}_{true}\mathbf{x}_{true}^T\right).\end{aligned} \tag{45}$$

From this, it is clear that $\mathbf{B}\mathbf{q} = \mathbf{0}$ for some $\mathbf{q} \in \mathbb{R}^n$, if and only if $\left(\mathbf{I} - \frac{1}{\ell}\mathbf{D}\mathbf{x}_{true}\mathbf{x}_{true}^T\right)\mathbf{q} = \mathbf{0}$. From the analysis subsequent to (41), it is found that $\mathbf{B}\mathbf{q} = \mathbf{0}$ if and only if $\mathbf{q} = \gamma\mathbf{D}\mathbf{x}_{true}$, for some $\gamma \in \mathbb{R}$.

Hence, its null-space is $\mathcal{N}[\mathbf{B}] = \text{span}\{\mathbf{D}\mathbf{x}_{true}\}$, and

$$\text{rank}\left[\mathbf{B}^T\right] = \text{rank}\left[\mathbf{B}\right] = n - 1. \tag{46}$$

On the other hand, combining (37) and (44) gives $\mathbf{x}_{true}^T\mathbf{D}\mathbf{B}\mathbf{H}^T\mathbf{W}\mathbf{v} = \mathbf{0}$. Since this must hold for all $\mathbf{v} \in \mathbb{R}^m$, it must in particular hold for all $\mathbf{v}$ of the form $\mathbf{v} = \mathbf{H}\mathbf{z}$, where $\mathbf{z} \in \mathbb{R}^n$. Consequently, $\mathbf{x}_{true}^T\mathbf{D}\mathbf{B}\overline{\mathbf{W}} = \mathbf{x}_{true}^T\mathbf{D}\mathbf{B}\mathbf{H}^T\mathbf{W}\mathbf{H} = \mathbf{0}$, which finally shows that $\mathbf{B}^T\mathbf{D}\mathbf{x}_{true} = \mathbf{0}$. As a result, $\mathbf{B}$ and $\mathbf{B}^T$ share the same null-space given by

$$\mathcal{N}[\mathbf{B}] = \mathcal{N}\left[\mathbf{B}^T\right] = \text{span}\{\mathbf{D}\mathbf{x}_{true}\}. \tag{47}$$

### B. Covariance Analysis

Suppose that the measurement error $\mathbf{v}$ in (29) is a random variable with mean and covariance $E[\mathbf{v}] = \mathbf{0}$ and $E\left[\mathbf{v}\mathbf{v}^T\right] = \mathbf{R}$ respectively, where $\mathbf{R} > 0$ is positive definite. Using (44), one obtains the mean and covariance of the estimation error

$$\delta\hat{\mathbf{x}} \triangleq E[\delta\mathbf{x}] = \mathbf{0}, \;\; \mathbf{P}_{\delta x} \triangleq E\left[\delta\mathbf{x}\delta\mathbf{x}^T\right] = \mathbf{B}\mathbf{H}^T\mathbf{W}\mathbf{R}\mathbf{W}\mathbf{H}\mathbf{B}^T. \tag{48}$$

Hence, the estimate $\mathbf{x}$ is unbiased. Furthermore, since $\mathbf{H}$ has full column rank, (46) and (47) give $\text{rank}[\mathbf{P}_{\delta x}] = \text{rank}\left[\mathbf{B}^T\right] = n-1$ and $\mathcal{N}[\mathbf{P}_{\delta x}] = \mathcal{N}\left[\mathbf{B}^T\right] = \text{span}\{\mathbf{D}\mathbf{x}_{true}\}$. Geometrically this makes sense, since it means that the estimate has no uncertainty in the direction perpendicular to the surface $\mathbf{x}_{true}^T\mathbf{D}\mathbf{x}_{true} = \ell$.

Often, the weighting matrix $\mathbf{W}$ in unconstrained least squares is chosen according to $\mathbf{W} = \mathbf{R}^{-1}$. In the unconstrained case, it can be shown that this choice minimizes trace $[\mathbf{P}_{\delta x}]$ [12]. It is not known if this still holds for the constrained case. However, with this choice of $\mathbf{W}$, the covariance in (48) becomes $\mathbf{P}_{\delta x} = \mathbf{B}\overline{\mathbf{W}}\mathbf{B}^T$, where $\overline{\mathbf{W}} = \mathbf{H}^T\mathbf{R}^{-1}\mathbf{H}$.

In practice, one does not know $\mathbf{x}_{true}$, so one cannot compute the covariance $\mathbf{P}_{\delta x}$ exactly. However, assuming that the computed least squares solution $\mathbf{x}$ is close to $\mathbf{x}_{true}$, the covariance can be approximated by replacing $\mathbf{x}_{true}$ by $\mathbf{x}$ in the expression for $\mathbf{B}$ in (45). The expression for $\mathbf{B}$ in (45) is independent of the value chosen for $\zeta \neq 0$. As such, it can be chosen to improve the conditioning of the matrix being inverted in the computation of $\mathbf{B}$.

## IV. Aerospace Applications

### A. Quadratically Constrained Kalman Filtering - Conventional Formulation

A norm-constrained Kalman filter was first presented in [4], and was later generalized in [5] to a Kalman filter with a quadratic constraint. This problem is revisited here, when the quadratic constraint has a positive semi-definite weight matrix. A more general language is used here compared to [4] and [5], to be able to incorporate such a constraint in other types of Kalman-like filter such as the unscented Kalman filter [15] and cubature Kalman filter [16], rather than just the standard and extended Kalman filters.

The constrained Kalman filters in [4] and [5] arise from enforcing the constraint in the correction step of the Kalman filter. As such, the focus here will also be on the correction step. To this end, let $\mathbf{x}_k$ be the state to be estimated at time $k$, $\hat{\mathbf{x}}_k^-$ a prior estimate of $\mathbf{x}_k$ resulting from the propagation step of a Kalman-type filter, $\mathbf{y}_k$ a measurement at time $k$, and $\hat{\mathbf{y}}_k^-$ a prior estimate of the measurement. Associated with the defined quantities are the covariances

$$\begin{gathered}\mathbf{P}_{xx,k}^- = E\left[(\mathbf{x}_k - \hat{\mathbf{x}}_k^-)(\mathbf{x}_k - \hat{\mathbf{x}}_k^-)^T\right], \ \ \mathbf{P}_{yy,k}^- = E\left[(\mathbf{y}_k - \hat{\mathbf{y}}_k^-)(\mathbf{y}_k - \hat{\mathbf{y}}_k^-)^T\right], \\ \mathbf{P}_{yx,k}^{-T} = \mathbf{P}_{xy,k}^- = E\left[(\mathbf{x}_k - \hat{\mathbf{x}}_k^-)(\mathbf{y}_k - \hat{\mathbf{y}}_k^-)^T\right].\end{gathered} \tag{49}$$

It is furthermore assumed that $\mathbf{P}_{xx,k}^-$ and $\mathbf{P}_{yy,k}^-$ are positive definite. The difference between the various types of Kalman filter is in how the prior estimates $\hat{\mathbf{x}}_k^-$, $\hat{\mathbf{y}}_k^-$ and the covariances $\mathbf{P}_{xx,k}^-$, $\mathbf{P}_{xy,k}^-$ and $\mathbf{P}_{yy,k}^-$ are obtained (or approximated).

All types of additive discrete-time Kalman filters have a correction step to obtain a posterior estimate $\hat{\mathbf{x}}_k$ of $\mathbf{x}_k$, of the form

$$\hat{\mathbf{x}}_k = \hat{\mathbf{x}}_k^- + \mathbf{K}_k(\mathbf{y}_k - \hat{\mathbf{y}}_k^-), \tag{50}$$

where $\mathbf{K}_k$ is a gain matrix. Defining the posterior covariance to be

$$\mathbf{P}_{xx,k} = E\left[(\mathbf{x}_k - \hat{\mathbf{x}}_k)(\mathbf{x}_k - \hat{\mathbf{x}}_k)^T\right], \tag{51}$$

using (49) and (50) yields

$$\mathbf{P}_{xx,k} = \mathbf{P}_{xx,k}^- - \mathbf{K}_k\mathbf{P}_{yx,k}^- - \mathbf{P}_{xy,k}^-\mathbf{K}_k^T + \mathbf{K}_k\mathbf{P}_{yy,k}^-\mathbf{K}_k^T. \tag{52}$$

The gain matrix $\mathbf{K}_k$ in (50) for any type of Kalman filter is then found to minimize trace$[\mathbf{P}_{xx,k}]$.

In the absence of any constraints, the minimizing gain is found to be

$$\mathbf{K}_{unc,k} = \mathbf{P}^{-}_{xy,k}\left(\mathbf{P}^{-}_{yy,k}\right)^{-1}. \tag{53}$$

Note that the subscript "unc" has been added in (53), since this is the optimal gain for an unconstrained posterior update.

The following quadratic constraint is now imposed on the posterior estimate

$$\hat{\mathbf{x}}_k^T \mathbf{D}_k \hat{\mathbf{x}}_k = \ell, \tag{54}$$

where $\mathbf{D}_k = \mathbf{D}_k^T \geq \mathbf{0}$ is symmetric and positive semi-definite, and $\ell > 0$. Note that this constraint covers the norm-constrained scenarios in [4], and is a special case of the general quadratic constraint considered in [5]. For the constrained Kalman filters, the gain $\mathbf{K}_k$ is found by solving the optimization problem:

**Problem 3**

Minimize

$$J(\mathbf{K}_k) = \text{trace}[\mathbf{P}_{xx,k}], \tag{55}$$

subject to (54) where $\mathbf{P}_{xx,k}$ is given by (52), and $\hat{\mathbf{x}}_k$ is given by (50).

To simplify future developments, the following transformation of variables is made

$$\mathbf{K}_k = \mathbf{K}_{unc,k} + \Delta\mathbf{K}_k, \tag{56}$$

where $\mathbf{K}_{unc,k}$ is the unconstrained optimal gain given by (53), and $\Delta\mathbf{K}_k$ becomes the new design variable. Substitution of (56) into (52) and making use of (53) leads to

$$\mathbf{P}_{xx,k} = \mathbf{P}_{xx,unc,k} + \Delta\mathbf{K}_k \mathbf{P}^{-}_{yy,k} \Delta\mathbf{K}_k^T, \tag{57}$$

where $\mathbf{P}_{xx,unc,k} = \mathbf{P}^{-}_{xx,k} - \mathbf{K}_{unc,k}\mathbf{P}^{-}_{yx,k} - \mathbf{P}^{-}_{xy,k}\mathbf{K}^T_{unc,k} + \mathbf{K}_{unc,k}\mathbf{P}^{-}_{yy,k}\mathbf{K}^T_{unc,k}$, is the posterior covariance resulting from an unconstrained optimal Kalman update. As such, Problem 3 is equivalent to

**Problem 4** Minimize

$$\hat{J}(\Delta\mathbf{K}_k) = \text{trace}[\Delta\mathbf{K}_k \mathbf{P}^{-}_{yy,k} \Delta\mathbf{K}_k^T], \tag{58}$$

subject to (54), where

$$\hat{\mathbf{x}}_k = \hat{\mathbf{x}}_{unc,k} + \Delta\mathbf{K}_k \mathbf{r}_k, \tag{59}$$

$\mathbf{r}_k = \mathbf{y}_k - \hat{\mathbf{y}}_k^-$ is the measurement residual, and $\hat{\mathbf{x}}_{unc,k} = \hat{\mathbf{x}}_k^- + \mathbf{K}_{unc,k}\mathbf{r}_k$ is the unconstrained posterior estimate.

The Lagrangian corresponding to Problem 4 is

$$L(\Delta\mathbf{K}_k, \lambda) = \hat{\mathbf{J}}(\Delta\mathbf{K}_k) + \lambda\left((\hat{\mathbf{x}}_{unc,k} + \Delta\mathbf{K}_k\mathbf{r}_k)^T\mathbf{D}_k(\hat{\mathbf{x}}_{unc,k} + \Delta\mathbf{K}_k\mathbf{r}_k) - \ell\right), \tag{60}$$

where $\lambda \in \mathbb{R}$ is the Lagrange multiplier. The necessary conditions for a solution to Problem 4 are then

$$\frac{\partial L}{\partial \Delta\mathbf{K}_k} = 2\left(\Delta\mathbf{K}_k\mathbf{P}_{yy,k}^- + \lambda\mathbf{D}_k(\hat{\mathbf{x}}_{unc,k} + \Delta\mathbf{K}_k\mathbf{r}_k)\mathbf{r}_k^T\right) = \mathbf{0}, \tag{61}$$

together with (54). It is shown in [5], that after appropriate transformations the necessary conditions in (54) and (61) result in an equation for the Lagrange multiplier $\lambda$, similar in form to the one in (12). It shown in both [4] and [5] that in the special case $\mathbf{D}_k = \mathbf{I}$ (norm-constrained case), an analytical solution for the minimizing Lagrange multiplier exists. However, in the more general case, the Lagrange multipliers must be solved for numerically, and the minimizing solution must be identified. It is clear from (61) that the Lagrange multiplier $\lambda$ depends on the unconstrained estimate $\hat{\mathbf{x}}_{unc,k}$. Therefore, at each time step, a new Lagrange multiplier must be found. Furthermore, the need to find the Lagrange multiplier and corresponding gain numerically does not yield any physical insight into the optimal constrained Kalman correction in general. A nice physical interpretation is currently only available in the literature in the norm-constrained case where the analytical solution for the Lagrange multiplier is used to show that the optimal constrained Kalman correction is simply an orthogonal projection of the unconstrained Kalman correction onto the constraint surface [4].

A different method of solution is proposed here, which will reduce the constrained Kalman correction to a problem identical in form to Problem 2, yielding physical insight into the constrained correction. This can then be solved using Algorithm 1 in this paper, without any need for computing Lagrange multipliers.

Before proceeding with the general case, the solution for the norm-constrained case where $\mathbf{D}_k = \mathbf{I}$ is recalled from [4], which is given by

$$\Delta\mathbf{K}_k = \frac{1}{r_k}\left(\frac{\sqrt{\ell}}{\|\hat{\mathbf{x}}_{unc,k}\|} - 1\right)\hat{\mathbf{x}}_{unc,k}\mathbf{r}_k^T\left(\mathbf{P}_{yy,k}^-\right)^{-1}, \tag{62}$$

where $r_k = \mathbf{r}_k^T \left( \mathbf{P}_{yy,k}^- \right)^{-1} \mathbf{r}_k$. The corresponding state estimate is

$$\hat{\mathbf{x}}_k = \frac{\sqrt{\ell}}{\|\hat{\mathbf{x}}_{unc,k}\|} \hat{\mathbf{x}}_{unc,k}, \tag{63}$$

which is simply the orthogonal projection of the unconstrained posterior estimate $\hat{\mathbf{x}}_{unc,k}$ onto the constraint surface $\hat{\mathbf{x}}_k^T \hat{\mathbf{x}}_k = \ell$. To see this, note that the outward pointing normal to the sphere at $\hat{\mathbf{x}}_k$ is given by $\nabla(\hat{\mathbf{x}}_k^T \hat{\mathbf{x}}_k - \ell) = 2\hat{\mathbf{x}}_k = 2(\sqrt{\ell}/\|\hat{\mathbf{x}}_{unc,k}\|)\hat{\mathbf{x}}_{unc,k}$. Consequently, $\hat{\mathbf{x}}_{unc,k} - \hat{\mathbf{x}}_k = (1 - (\sqrt{\ell}/\|\hat{\mathbf{x}}_{unc,k}\|))\hat{\mathbf{x}}_{unc,k}$, which is parallel to the unit normal at $\hat{\mathbf{x}}_k$, and therefore is orthogonal to the tangent plane at $\hat{\mathbf{x}}_k$. A similar interpretation is now obtained for the more general case when $\mathbf{D}_k$ is simply required to be positive semi-definite.

To make this problem tractable, a further restriction on the gain $\Delta\mathbf{K}_k$ is made. Inspired by (62) in the norm-constrained case ($\mathbf{D}_k = \mathbf{I}$), the gain is restricted to have the form

$$\Delta\mathbf{K}_k = \frac{1}{r_k} \tilde{\mathbf{k}}_k \mathbf{r}_k^T \left( \mathbf{P}_{yy,k}^- \right)^{-1}, \tag{64}$$

where $\tilde{\mathbf{k}}_k$ is the new design variable. As will become apparent shortly, this restriction does not result in any loss of generality. In making this restriction, the constrained correction in (59) becomes

$$\hat{\mathbf{x}}_k = \hat{\mathbf{x}}_{unc,k} + \tilde{\mathbf{k}}_k, \tag{65}$$

and the cost function (58) becomes

$$\hat{J}(\Delta\mathbf{K}_k(\tilde{\mathbf{k}}_k)) = \tilde{\mathbf{k}}_k^T \tilde{\mathbf{k}}_k / r_k. \tag{66}$$

Noting the relationship between $\tilde{\mathbf{k}}_k$ and the constrained posterior estimate $\hat{\mathbf{x}}_k$ in (65), with the gain restriction in (64), the minimization problem in Problem 4 becomes equivalent to the orthogonal projection problem:

**Problem 5** Minimize

$$\tilde{J}(\hat{\mathbf{x}}_k) = \frac{1}{2}(\hat{\mathbf{x}}_{unc,k} - \hat{\mathbf{x}}_k)^T (\hat{\mathbf{x}}_{unc,k} - \hat{\mathbf{x}}_k), \tag{67}$$

subject to $\hat{\mathbf{x}}_k^T \mathbf{D}_k \hat{\mathbf{x}}_k = \ell$.

Thus, with the gain restriction in (64), the constrained posterior estimate is simply the orthogonal projection of the unconstrained posterior estimate $\hat{\mathbf{x}}_{unc,k}$ onto the set defined by the constraint

$\hat{\mathbf{x}}_k^T \mathbf{D}_k \hat{\mathbf{x}}_k = \ell$. This problem is readily solved using Algorithm 1. Furthermore, having found $\hat{\mathbf{x}}_k$, the corresponding gain $\Delta\mathbf{K}_k$ (which is needed for the covariance-like update in (57)) is readily computed by substitution of (65) into (64) and is given by

$$\Delta\mathbf{K}_k = \frac{1}{r_k}(\hat{\mathbf{x}}_k - \hat{\mathbf{x}}_{unc,k})\mathbf{r}_k^T \left(\mathbf{P}_{yy,k}^-\right)^{-1}. \tag{68}$$

It is now shown that the gain (68) resulting from the solution of Problem 5 satisfies the necessary condition for Problem 4. First, applying the necessary condition for Problem 2 in (7) to Problem 5, the necessary condition for a solution to Problem 5 is

$$\hat{\mathbf{x}}_{unc,k} - \hat{\mathbf{x}}_k = \bar{\lambda}\mathbf{D}_k\hat{\mathbf{x}}_k, \tag{69}$$

where $\bar{\lambda} \in \mathbb{R}$ is a Lagrange multiplier. Consequently, it can be shown that the gain (68) satisfies the necessary condition for Problem 4 in (61) with Lagrange multiplier $\lambda = \bar{\lambda}/r$.

A geometric interpretation has now been obtained for the quadratically constrained Kalman filter problem defined in Problem 3, when the gain is restricted to the form in (64). Namely, the optimal constrained posterior state estimate is simply the orthogonal projection of the optimal unconstrained posterior estimate onto the set defined by (54). Furthermore, as already mentioned this problem can be solved using Algorithm 1, without the need to compute any Lagrange multiplier.

It was stated in the introduction of [5] that when the weight matrix $\mathbf{D}$ is positive-definite, the constrained Kalman filter problem in Problem 3 could be equivalently solved by a coordinate transformation to yield a norm constraint, and then apply the norm-constrained solution. It is now shown that this is generally not true. To this end, let $\mathbf{V}_k$ be a matrix such that $\mathbf{V}_k^T\mathbf{V}_k = \mathbf{D}_k$, and consider the coordinate transformation $\hat{\mathbf{z}}_k = \mathbf{V}_k\hat{\mathbf{x}}_k$. Then, in the transformed coordinates, the constraint in (54) becomes $\hat{\mathbf{z}}_k^T\hat{\mathbf{z}}_k = \ell$. It can be shown that the unconstrained posterior estimate in transformed coordinates is given by $\hat{\mathbf{z}}_{unc,k} = \mathbf{V}_k\hat{\mathbf{x}}_{unc,k}$. From (63), the optimal constrained posterior estimate for the transformed coordinates (in the sense of minimizing trace$[\mathbf{P}_{zz,k}]$, where $\mathbf{P}_{zz,k} = \mathbf{V}_k\mathbf{P}_{xx,k}\mathbf{V}_k^T$) is given by $\hat{\mathbf{z}}_k = (\sqrt{\ell}/\|\hat{\mathbf{z}}_{unc,k}\|)\hat{\mathbf{z}}_{unc,k}$, with corresponding gain $\Delta\mathbf{K}_{z,k} = \frac{1}{r_k}\left((\sqrt{\ell}/\|\hat{\mathbf{z}}_{unc,k}\|) - 1\right)\hat{\mathbf{z}}_{unc,k}\mathbf{r}_k^T\left(\mathbf{P}_{yy,k}^-\right)^{-1}$. However, transforming this back to the original coordinates gives

$$\hat{\mathbf{x}}_k = \frac{\sqrt{\ell}}{\sqrt{\hat{\mathbf{x}}_{unc,k}^T \mathbf{D}_k \mathbf{x}_{unc,k}}} \hat{\mathbf{x}}_{unc,k}, \tag{70}$$

together with gain

$$\Delta\mathbf{K}_k = \frac{1}{r_k} \left( \frac{\sqrt{\ell}}{\sqrt{\hat{\mathbf{x}}_{unc,k}^T \mathbf{D}_k \mathbf{x}_{unc,k}}} - 1 \right) \hat{\mathbf{x}}_{unc,k} \mathbf{r}_k^T \left( \mathbf{P}_{yy,k}^- \right)^{-1}. \tag{71}$$

The gain in (71) has the form in (64). However, the state estimate in (70) is just a scaling of the unconstrained estimate $\hat{\mathbf{x}}_{unc,k}$ onto the constraint surface, which in general is not an orthogonal projection, unless $\hat{\mathbf{x}}_{unc,k}$ is parallel to a principal axis of the constraint ellipsoid. As such, the gain in (68) generally yields a lower value of the cost in (58) than the gain in (71). Hence, the transformed problem generally does not solve Problem 5, and consequently does not solve the original problem. This demonstrates that unlike the unconstrained case, the optimal constrained Kalman update is specific to the coordinate representation of the system.

With continuous-discrete Kalman filter implementations, it is often of interest to obtain optimal estimates in between measurement updates, particularly when measurement updates are infrequent. For a quadratically constrained extended Kalman filter, the state estimate in between updates would typically satisfy the constraint provided that the constraint is a constant of motion of the nonlinear dynamic equation being propagated. However, for sigma point type Kalman filters such as the unscented Kalman filter and the cubature Kalman filter, between measurement updates, the state estimate is simply a weighted average of the sigma points being propagated in between measurement updates, and will typically violate the constraint. In this case, let $\hat{\mathbf{x}}^-(t)$ be the state estimate between measurement updates ($t_k < t < t_{k+1}$, where $t_k$ are measurement times), together with covariance $\mathbf{P}_{xx}^-(t)$. A suitable constrained estimate can then be obtained by minimizing $J(\hat{\mathbf{x}}(t)) = \frac{1}{2}(\hat{\mathbf{x}}^-(t) - \hat{\mathbf{x}}(t))^T (\mathbf{P}_{xx}^-(t))^{-1} (\hat{\mathbf{x}}^-(t) - \hat{\mathbf{x}}(t))$, subject to $\hat{\mathbf{x}}(t)^T \mathbf{D} \hat{\mathbf{x}}(t) = \ell$. This is readily solved using Algorithm 1 in this paper.

### B. Quadratically Constrained Kalman Filtering - Alternative Formulation

The constrained Kalman filter from [4, 5], and the previous section, all assume that the covariance update satisfies (52). However, as seen in (68), the gain $\mathbf{K}_k = \mathbf{K}_{unc,k} + \Delta\mathbf{K}_k$ in the constrained filter depends on the measurement residual $\mathbf{r}_k = \mathbf{y}_k - \hat{\mathbf{y}}_k^-$. Consequently, the gain $\mathbf{K}_k$ is itself a random variable, and therefore the covariance does not satisfy (52). That is, $\mathbf{P}_{xx,k}$ can no longer be interpreted as the covariance of the state estimate $\hat{\mathbf{x}}_k$. To obtain a constrained Kalman filter with a more consistent statistical interpretation, an alternative approach is proposed here.

Unlike the previous section, this section will be restricted to the standard and extended versions of the Kalman filter [12]. As in the previous section, the difference between the unconstrained and constrained versions of these filters is exclusively in the correction step (the prediction step remains identical).

*1. Linear Measurements*

First consider the standard unconstrained Kalman filter. As in the previous section, for simplicity of notation time-dependence of variables is dropped. A prior estimate $\hat{\mathbf{x}}_k^-$ of the state $\mathbf{x}_k$ is assumed available, with positive semi-definite covariance $\mathbf{P}_{xx,k}^- = E\{(\mathbf{x}_k - \hat{\mathbf{x}}_k^-)(\mathbf{x}_k - \hat{\mathbf{x}}_k^-)^T\}$, together with a measurement of the form

$$\mathbf{y}_k = \mathbf{H}_k \mathbf{x}_k + \mathbf{v}_k, \tag{72}$$

where $\mathbf{v}_k$ is a zero-mean random variable with positive-definite covariance $\mathbf{R}_k = E\{\mathbf{v}_k \mathbf{v}_k^T\}$. It is furthermore assumed that $\tilde{\mathbf{x}}_k^- = \mathbf{x}_k - \hat{\mathbf{x}}_k^-$ and $\mathbf{v}_k$ are uncorrelated (which is satisfied for the standard Kalman filter). In the unconstrained case, the covariance $\mathbf{P}_{xx,k}^-$ is actually positive-definite, and the state update in the Kalman filter can be obtained by solving the minimization problem [17]

$$\hat{\mathbf{x}}_k = \text{argmin}\left\{\frac{1}{2}(\hat{\mathbf{x}}_k^- - \mathbf{x}_k)^T(\mathbf{P}_{xx,k}^-)^{-1}(\hat{\mathbf{x}}_k^- - \mathbf{x}_k) + \frac{1}{2}(\mathbf{y}_k - \mathbf{H}_k\mathbf{x}_k)^T\mathbf{R}_k^{-1}(\mathbf{y}_k - \mathbf{H}_k\mathbf{x}_k)\right\}. \tag{73}$$

This can be concisely written as the standard weighted least squares problem

$$\hat{\mathbf{x}}_k = \text{argmin}\frac{1}{2}(\bar{\mathbf{y}}_k - \bar{\mathbf{H}}_k\mathbf{x}_k)^T\bar{\mathbf{R}}_k^{-1}(\bar{\mathbf{y}}_k - \bar{\mathbf{H}}_k\mathbf{x}_k), \tag{74}$$

where

$$\bar{\mathbf{y}}_k = \begin{bmatrix} \hat{\mathbf{x}}_k^- \\ \mathbf{y}_k \end{bmatrix}, \quad \bar{\mathbf{H}}_k = \begin{bmatrix} \mathbf{I} \\ \mathbf{H}_k \end{bmatrix}, \quad \bar{\mathbf{R}}_k = E\left\{\begin{bmatrix} \tilde{\mathbf{x}}_k^- \\ \mathbf{v}_k \end{bmatrix}\begin{bmatrix} (\tilde{\mathbf{x}}_k^-)^T & \mathbf{v}_k^T \end{bmatrix}\right\} = \text{diag}\{\mathbf{P}_{xx,k}^-, \mathbf{R}_k\}.$$

For the constrained Kalman filter, with constraint given in (54) with $\mathbf{D}_k$ positive semi-definite as before, it is proposed here that the state estimate correction step be given by replacing the unconstrained minimization problem in (74) by the constrained minimization problem

$$\hat{\mathbf{x}}_k = \text{argmin}\frac{1}{2}(\bar{\mathbf{y}}_k - \bar{\mathbf{H}}_k\mathbf{x}_k)^T\tilde{\mathbf{R}}_k^{-1}(\bar{\mathbf{y}}_k - \bar{\mathbf{H}}_k\mathbf{x}_k), \tag{75}$$

subject to (54), where

$$\tilde{\mathbf{R}}_k = \text{diag}\{\mathbf{P}_{xx,k}^{-} + \delta\mathbf{I}, \mathbf{R}_k\}, \tag{76}$$

with $\delta > 0$ a user defined parameter ensuring that $\tilde{\mathbf{R}}_k$ is invertible, since now $\mathbf{P}_{xx,k}^{-}$ is allowed to be singular.

As shown in Sections II and III of this paper, the state estimate correction step then decomposes into three stages:

1. Solve the unconstrained optimization problem

$$\hat{\mathbf{x}}_{unc,k} = \text{argmin}\frac{1}{2}(\bar{\mathbf{y}}_k - \bar{\mathbf{H}}_k\mathbf{x}_k)^T\tilde{\mathbf{R}}_k^{-1}(\bar{\mathbf{y}}_k - \bar{\mathbf{H}}_k\mathbf{x}_k), \tag{77}$$

2. Solve the subsequent constrained optimization problem

$$\hat{\mathbf{x}}_k = \text{argmin}\frac{1}{2}(\hat{\mathbf{x}}_{unc,k} - \mathbf{x}_k)^T\overline{\mathbf{W}}_k(\hat{\mathbf{x}}_{unc,k} - \mathbf{x}_k), \tag{78}$$

   subject to $\hat{\mathbf{x}}_k^T\mathbf{D}_k\hat{\mathbf{x}}_k = \ell$, where $\overline{\mathbf{W}}_k = \bar{\mathbf{H}}_k^T\tilde{\mathbf{R}}_k^{-1}\bar{\mathbf{H}}_k$.

3. The covariance of the corrected state estimate can now be approximated using (48) as

$$\mathbf{P}_{xx,k} = E\{(\mathbf{x}_k - \hat{\mathbf{x}}_k)(\mathbf{x}_k - \hat{\mathbf{x}}_k)^T\} = \mathbf{B}_k\bar{\mathbf{H}}_k^T\tilde{\mathbf{R}}_k^{-1}\bar{\mathbf{R}}_k\tilde{\mathbf{R}}_k^{-1}\bar{\mathbf{H}}_k\mathbf{B}_k^T, \tag{79}$$

   where

$$\begin{aligned}\mathbf{B}_k &= \left(\overline{\mathbf{W}}_k\left(\mathbf{I} - \frac{1}{\ell}\mathbf{D}_k\hat{\mathbf{x}}_k\hat{\mathbf{x}}_k^T\right)^T\left(\mathbf{I} - \frac{1}{\ell}\mathbf{D}_k\hat{\mathbf{x}}_k\hat{\mathbf{x}}_k^T\right)\overline{\mathbf{W}}_k + \zeta^2\mathbf{D}_k\hat{\mathbf{x}}_k\hat{\mathbf{x}}_k^T\mathbf{D}_k\right)^{-1} \\ &\quad \times\overline{\mathbf{W}}_k\left(\mathbf{I} - \frac{1}{\ell}\mathbf{D}_k\hat{\mathbf{x}}_k\hat{\mathbf{x}}_k^T\right)^T\left(\mathbf{I} - \frac{1}{\ell}\mathbf{D}_k\hat{\mathbf{x}}_k\hat{\mathbf{x}}_k^T\right),\end{aligned} \tag{80}$$

   and $\zeta > 0$ is a user defined parameter.

Thus, for the proposed constrained Kalman filter, the state estimate correction step is given by the pair of minimization problems in (77) and (78), while the covariance update is given by (79).

As discussed in Section III.B, the covariance $\mathbf{P}_{xx,k}$ in (79) is in fact singular, which is reflective of the fact that the estimate $\hat{\mathbf{x}}$ is constrained to the set defined by $\hat{\mathbf{x}}_k^T\mathbf{D}_k\hat{\mathbf{x}}_k = \ell$. This is the reason $\mathbf{P}_{xx,k}^{-}$ is allowed to be singular. There is no guarantee it will become non-singular after the propagation step of the Kalman filter. An interesting question is what the best choice of $\delta > 0$ in (76) would in terms of leading to the smallest posterior covariance $\mathbf{P}_{xx,k}$ in (79). This is not something that has been investigated thus far. However, it should be noted that the resulting covariance expression in equation (79) is derived from the covariance expression in equation (48), which is valid for any choice of positive-definite weighting matrix in the least squares problem. Therefore, regardless of choice of $\delta > 0$ in (76), the resulting posterior covariance $\mathbf{P}_{xx,k}$ is correct for the obtained posterior estimate $\hat{\mathbf{x}}_k$ in (78), even if it isn't minimum variance.

### *2. Extension to Nonlinear Measurements*

Consider the case where the measurement equation (72) is replaced by the nonlinear measurement equation

$$\mathbf{y}_k = \mathbf{h}_k(\mathbf{x}_k) + \mathbf{v}_k, \tag{81}$$

where $\mathbf{h}_k(\mathbf{x}_k)$ is assumed to be differentiable. Correspondingly, the minimization problem (75) is replaced by

$$\hat{\mathbf{x}}_k = \text{argmin}\frac{1}{2}(\bar{\mathbf{y}}_k - \bar{\mathbf{h}}_k(\mathbf{x}_k))^T\tilde{\mathbf{R}}_k^{-1}(\bar{\mathbf{y}}_k - \bar{\mathbf{h}}_k(\mathbf{x}_k)), \tag{82}$$

subject to (54), where

$$\bar{\mathbf{h}}_k(\mathbf{x}_k) = \begin{bmatrix} \mathbf{x}_k \\ \mathbf{h}_k(\mathbf{x}_k) \end{bmatrix},$$

and $\tilde{\mathbf{R}}_k$ is given by (76) as before. Similar to the derivation of the unconstrained extended Kalman filter, a reference state $\bar{\mathbf{x}}_k$ is assumed available such that $\mathbf{h}_k(\mathbf{x}_k)$ is well approximated by $\mathbf{h}_k(\mathbf{x}_k) = \mathbf{h}_k(\bar{\mathbf{x}}_k) + (\partial\mathbf{h}_k(\bar{x}_k)/\partial\mathbf{x}_k)(\mathbf{x}_k - \bar{\mathbf{x}}_k)$. As such, the minimization problem in (82) becomes

$$\hat{\mathbf{x}}_k = \text{argmin}\frac{1}{2}(\tilde{\mathbf{y}}_k - \tilde{\mathbf{H}}_k(\bar{\mathbf{x}}_k)\mathbf{x}_k)^T\tilde{\mathbf{R}}_k^{-1}(\tilde{\mathbf{y}}_k - \tilde{\mathbf{H}}_k(\bar{\mathbf{x}}_k)\mathbf{x}_k), \tag{83}$$

subject to (54), where

$$\tilde{\mathbf{y}}_k = \begin{bmatrix} \hat{\mathbf{x}}_k^{-} \\ \mathbf{y}_k - \mathbf{h}_k(\bar{\mathbf{x}}_k) + (\partial\mathbf{h}_k(\bar{\mathbf{x}}_k)/\partial\mathbf{x}_k)\bar{\mathbf{x}}_k \end{bmatrix}, \quad \tilde{\mathbf{H}}_k(\bar{\mathbf{x}}_k) = \frac{\partial\bar{\mathbf{h}}_k}{\partial\mathbf{x}_k}(\bar{\mathbf{x}}_k) = \begin{bmatrix} \mathbf{I} \\ (\partial\mathbf{h}_k(\bar{\mathbf{x}}_k)/\partial\mathbf{x}_k) \end{bmatrix}. \tag{84}$$

Problem (83) is a linear problem of the same form as that in (75), which can be solved by the three step approach described in (77) and (78), with resulting covariance in (79). Finally, an appropriate reference state $\bar{\mathbf{x}}_k$ is chosen. Similar to the derivation of the unconstrained extended Kalman filter, for given realizations of the prior estimate $\hat{\mathbf{x}}_k^-$ and measurement $\mathbf{y}_k$ respectively, let $\bar{\mathbf{x}}_k$ be the minimizing solution to the unconstrained minimization problem

$$\bar{\mathbf{x}}_k = \text{argmin}\frac{1}{2}(\bar{\mathbf{y}}_k - \bar{\mathbf{h}}_k(\mathbf{x}_k))^T\tilde{\mathbf{R}}_k^{-1}(\bar{\mathbf{y}}_k - \bar{\mathbf{h}}_k(\mathbf{x}_k)). \tag{85}$$

The corresponding necessary condition can be shown to be $\tilde{\mathbf{H}}_k^T(\bar{\mathbf{x}}_k)\tilde{\mathbf{R}}_k^{-1}(\bar{\mathbf{y}}_k - \bar{\mathbf{h}}_k(\bar{\mathbf{x}}_k)) = \mathbf{0}$. Substituting this into (83) leads to

$$\hat{\mathbf{x}}_k = \text{argmin}\frac{1}{2}(\bar{\mathbf{x}}_k - \mathbf{x}_k)^T\tilde{\mathbf{H}}_k^T(\bar{\mathbf{x}}_k)\tilde{\mathbf{R}}_k^{-1}\tilde{\mathbf{H}}_k(\bar{\mathbf{x}}_k)(\bar{\mathbf{x}}_k - \mathbf{x}_k), \tag{86}$$

subject to (54). Finally, denoting $\hat{\mathbf{x}}_{unc,k} = \bar{\mathbf{x}}_k$ and applying (77), (78), and (79) to (83) gives the three-step constrained state estimate update:

1. Solve the unconstrained optimization problem

$$\hat{\mathbf{x}}_{unc,k} = \text{argmin}\frac{1}{2}(\bar{\mathbf{y}}_k - \bar{\mathbf{h}}_k(\mathbf{x}_k))^T\tilde{\mathbf{R}}_k^{-1}(\bar{\mathbf{y}}_k - \bar{\mathbf{h}}_k(\mathbf{x}_k)), \tag{87}$$

2. Solve the subsequent constrained optimization problem

$$\hat{\mathbf{x}}_k = \text{argmin}\frac{1}{2}(\hat{\mathbf{x}}_{unc,k} - \mathbf{x}_k)^T\overline{\mathbf{W}}_k(\hat{\mathbf{x}}_{unc,k} - \mathbf{x}_k), \tag{88}$$

subject to $\hat{\mathbf{x}}_k^T\mathbf{D}_k\hat{\mathbf{x}}_k = \ell$, where $\overline{\mathbf{W}}_k = \tilde{\mathbf{H}}_k^T(\hat{\mathbf{x}}_{unc,k})\tilde{\mathbf{R}}_k^{-1}\tilde{\mathbf{H}}_k(\hat{\mathbf{x}}_{unc,k})$, and $\tilde{\mathbf{H}}_k$ is given in (84).

3. The covariance of the corrected state estimate is now approximated as

$$\mathbf{P}_{xx,k} = \mathbf{B}_k\tilde{\mathbf{H}}_k^T(\hat{\mathbf{x}}_{unc,k})\tilde{\mathbf{R}}_k^{-1}\bar{\mathbf{R}}_k\tilde{\mathbf{R}}_k^{-1}\tilde{\mathbf{H}}_k(\hat{\mathbf{x}}_{unc,k})\mathbf{B}_k^T, \tag{89}$$

where $\mathbf{B}_k$ is given by (80).

### *3. Numerical Example*

A spacecraft attitude estimation example is used to compare the proposed constrained filter with the norm-constrained filter [4]. The actual implementation of the norm-constrained filter is as in [18], where the attitude update step is truly additive, rather than in [4], where the attitude update step is multiplicative.

The spacecraft is in a circular Keplerian orbit with altitude 450 $km$, inclination $87^o$, right ascension of ascending node $100^o$, and initial argument of latitude $0^o$ at epoch noon, 1 Feb, 2010. The only measurement available is from a magnetometer providing a measurement of the Earth's magnetic field vector in spacecraft body coordinates with a sample period of 10 seconds. The Earth's magnetic field is obtained using a $10^{th}$ order International Geomagnetic Reference Field (IGRF) model [19]. The spacecraft is freely tumbling and is acted upon by a gravity gradient torque [20] and a magnetic disturbance torque [20] due to a residual dipole moment given in spacecraft body coordinates by $\mathbf{m}_b = [0.1\ 0.1\ 0.1]^T$ A/m$^2$. However, these torques are not included in the filters' process models. The spacecraft inertia matrix is given by $\mathbf{J} = \text{diag}\{27, 17, 25\}$ kg·m$^2$. The spacecraft attitude is represented by the unit quaternion $\mathbf{q} = [\boldsymbol{\epsilon}^T\ \eta]^T$, where $\boldsymbol{\epsilon} \in \mathbb{R}^3$ and $\eta \in \mathbb{R}$ are the vector and scalar parts of the quaternion, respectively. The quaternion satisfies the unit length constraint $\mathbf{q}^T\mathbf{q} = 1$.

The measurement model used is $\mathbf{y}(t_k) = \mathbf{C}_{bI}(\mathbf{q}(t_k))\hat{\mathbf{b}}_I(t_k) + \mathbf{v}_k$, where $\hat{\mathbf{b}}_I$ is the normalized Earth's magnetic field vector in inertial coordinates, $\mathbf{v}_k$ is zero-mean measurement noise with covariance $E\{\mathbf{v}_k\mathbf{v}_j^T\} = \delta_{kj} r \mathbf{I}_3$ with $r = 10^{-4}$, where $\delta_{kj}$ is the discrete delta function, and $\mathbf{C}_{bI}(\mathbf{q}) = (\eta^2 - \boldsymbol{\epsilon}^T\boldsymbol{\epsilon})\mathbf{I} + 2\boldsymbol{\epsilon}\boldsymbol{\epsilon}^T - 2\eta\boldsymbol{\epsilon}^\times$, is the rotation matrix transforming vector representations from inertial to spacecraft body coordinates. The notation $\mathbf{a}^\times \in \mathbb{R}^{3\times3}$ refers to the skew-symmetric matrix (i.e., $\mathbf{a}^\times = -\mathbf{a}^{\times^T}$) that realizes a cross product, that is, given $\mathbf{a} \in \mathbb{R}^3$ and $\mathbf{b} \in \mathbb{R}^3$ their cross product is $\mathbf{a}^\times\mathbf{b} \in \mathbb{R}^3$.

The process model assumed in the filters is given by the quaternion kinematics together with Euler's equations $\dot{\boldsymbol{\epsilon}} = (1/2)(\boldsymbol{\epsilon}^\times + \eta\mathbf{I})\boldsymbol{\omega}$, $\dot{\eta} = -(1/2)\boldsymbol{\epsilon}^T\boldsymbol{\omega}$ and $\dot{\boldsymbol{\omega}} = -\mathbf{J}^{-1}\boldsymbol{\omega}^\times\mathbf{J}\boldsymbol{\omega} + \mathbf{J}^{-1}\mathbf{w}$, where $\boldsymbol{\omega} \in \mathbb{R}^3$ is the spacecraft angular velocity vector resolved in the spacecraft body frame, and $\mathbf{w} \in \mathbb{R}^3$ is a zero-mean noise process with covariance $E\{\mathbf{w}(t)\mathbf{w}(\tau)^T\} = \delta(t-\tau)q\mathbf{I}_3$ with $q = 2 \times 10^{-7}$, where $\delta(t-\tau)$ is the Dirac delta function. The state to be estimated is then $\mathbf{x} = [\mathbf{q}^T\ \boldsymbol{\omega}^T]^T$. Furthermore, due to the unit length constraint on the quaternion, the state estimate must satisfy $\hat{\mathbf{x}}^T\mathbf{D}\hat{\mathbf{x}} = \ell$, where $\mathbf{D} = \text{diag}\{\mathbf{I}_4, \mathbf{0}_{3\times3}\}$ and $\ell = 1$. Note that reference [4] also presents a norm-constrained filter where only a part of the state vector has a norm constraint, such as is the case in this example.

The true spacecraft initial conditions are $\mathbf{q}(0) = [\sin(1/2)/\sqrt(3)[1\ 1\ 1]\cos(1/2)]^T$, $\boldsymbol{\omega}(0) =$

$[0.05\ \ -0.05\ 0.05]^T$ rad/s. Both filters are given the initial conditions $\hat{\mathbf{q}}(0) = [0\ 0\ 0\ 1]^T$, $\hat{\boldsymbol{\omega}}(0) = [0\ 0\ 0]^T$ rad/s. The filters are both given the initial covariance $\mathbf{P}_{xx}(0) = 0.1\mathbf{I}_7$. Finally, for the proposed filter in this section, $\tilde{\mathbf{R}}$ in (76) was formed with $\delta = 10^{-5}\tanh^2(\|\mathbf{y}(t_k) - \mathbf{C}_{bI}(\hat{\mathbf{q}}^-(t_k))\hat{\mathbf{b}}_I(t_k)\|)$. It was found that this improved filter convergence. By using this type of weight, it ensures that the prior estimate is weighted less in the minimization problem in (82) when the measurement residual is large, which might not otherwise occur, since $\mathbf{P}_{xx}^-(t_k)$ may be close to singular due to the posterior covariance from the previous time step given in (89).

Finally, in this example, the unconstrained minimization in (87) is solved using the Levenberg-Marquardt method [12], and the constrained minimization in (88) is solved using Algorithm 1 in this paper, using the line-search method from [13] to obtain step lengths satisfying the Wolfe conditions in (27) and (28) with parameters $c_1 = 10^{-4}$, $c_2 = 0.9$. To concisely present the attitude and angular velocity estimation errors, the attitude estimation error is represented by the principal angle of rotation from the true attitude to the estimated attitude, given by [20] $\phi = \cos^{-1}\big((\text{trace}[\mathbf{C}_{bI}(\hat{\mathbf{q}})\mathbf{C}_{bI}^T(\mathbf{q})] - 1)/2\big)$. The angular velocity estimation error is represented by $\tilde{\omega} = \|\hat{\boldsymbol{\omega}} - \boldsymbol{\omega}\|_2$.

Figures 1 and 2 show the resulting attitude and angular velocity estimation errors respectively, for the two filters. Table 1 shows the average estimation errors once the filters have converged (from 0.5 orbits onwards). It can be seen that the proposed constrained filter slightly outperforms the conventional norm-constrained filter, but overall the performance is very similar. This actually demonstrates the efficacy of the conventional norm-constrained Kalman filter in that it performs very similarly to the more statistically consistent filter proposed here. On the other hand, the benefit of using the proposed constrained filter is that $\mathbf{P}_{xx}$ retains the statistical interpretation of being the covariance of the constrained estimate $\hat{\mathbf{x}}$, which is not the case for the conventional norm-constrained filter.

## V. Conclusions

A new method has been presented for solving quadratically constrained least squares problems without the need for determining Lagrange multipliers, with an algorithm that searches directly

**Table 1 Mean Estimation Errors (Steady-State)**

| Estimation Scheme | $\phi_{av}$ (deg) | $\tilde{\omega}_{av}$ (deg/s) |
| --- | --- | --- |
| Conventional Norm-Constrained [4, 18] | 0.4576 | 0.0187 |
| Proposed Norm-Constrained | 0.4242 | 0.0186 |

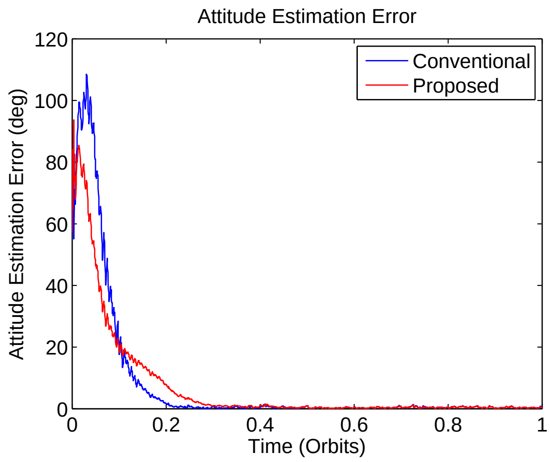


**Fig. 1 Attitude Estimation Error Comparison**

over the constraint set. Global convergence of the algorithm to the set of stationary points has been rigorously proven. A covariance expression has been obtained for the constrained optimal solution. Two aerospace applications of the new method have been presented. The first application is the generalization of a norm-constrained Kalman filter from the literature to a quadratically constrained Kalman filter. It is shown that the optimal quadratically constrained update is simply an orthogonal projection of the optimal unconstrained update onto the constraint set, which is readily performed using the method presented in this paper. The second application is a new quadratically constrained Kalman filter using the covariance expression developed in this paper, yielding a statistically more consistent filter compared with the norm-constrained Kalman filter from the literature. The new filter is demonstrated numerically with a spacecraft attitude estimation example.

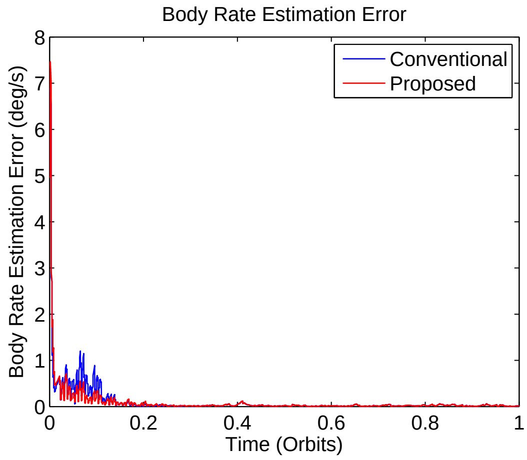


**Fig. 2 Body-Rate Estimation Error Comparison**

## APPENDIX A: MATHEMATICAL PROOFS

### 1. Proof of Proposition 1

It is first assumed that $0 < \beta_k < 1$. From (22), it can be seen that $d\phi_k(0)/d\alpha = 0$ if and only if both

$$\left(\mathbf{I} - \frac{\mathbf{D}\mathbf{x}_k\mathbf{x}_k^T\mathbf{D}}{\|\mathbf{D}\mathbf{x}_k\|_2^2}\right)\overline{\mathbf{W}}\left(\mathbf{x}_{unc} - \mathbf{x}_k\right) = 0, \text{ and } \left(\overline{\mathbf{W}} - \frac{\mathbf{D}\mathbf{x}_k\mathbf{x}_k^T\mathbf{D}}{\mathbf{x}_k^T\mathbf{D}\overline{\mathbf{W}}^{-1}\mathbf{D}\mathbf{x}_k}\right)\left(\mathbf{x}_{unc} - \mathbf{x}_k\right) = 0. \tag{A1}$$

It can be shown that the first equality in (A1) is satisfied if and only if $\overline{\mathbf{W}}(\mathbf{x}_{unc} - \mathbf{x}_k) = \lambda\mathbf{D}\mathbf{x}_k$, for some $\lambda \in \mathbb{R}$, which is precisely the necessary condition in (7). To obtain the necessary and sufficient condition for the second equality in (A1), note that $\mathbf{x}_k^T\mathbf{D}\mathbf{x}_k = \ell > 0$, and therefore $\mathbf{D}\mathbf{x}_k \neq \mathbf{0}$. Now, let $\mathbf{z}_1, ..., \mathbf{z}_{n-1} \in \mathbb{R}^n$ extend $\mathbf{D}\mathbf{x}_k$ to an orthogonal basis for $\mathbb{R}^n$. Then, $\overline{\mathbf{W}}^{-1}\mathbf{D}\mathbf{x}_k, \mathbf{z}_1, ..., \mathbf{z}_{n-1}$ also forms a basis for $\mathbb{R}^n$. To see why, consider $\lambda\overline{\mathbf{W}}^{-1}\mathbf{D}\mathbf{x}_k + \sum_{i=1}^{n-1}\alpha_i\mathbf{z}_i = \mathbf{0}$, for some $\lambda, \alpha_1, ..., \alpha_{n-1} \in \mathbb{R}$. Premultiplying by $\mathbf{x}_k^T\mathbf{D}$ gives $\lambda\mathbf{x}_k^T\mathbf{D}\overline{\mathbf{W}}^{-1}\mathbf{D}\mathbf{x}_k = 0$. Consequently $\lambda = 0$, and since $\mathbf{z}_1, ..., \mathbf{z}_{n-1}$ are linearly independent, this implies that $\alpha_1 = \cdots = \alpha_{n-1} = 0$ also. Therefore, one can write

$$\mathbf{x}_{unc} - \mathbf{x}_k = \lambda\overline{\mathbf{W}}^{-1}\mathbf{D}\mathbf{x}_k + \sum_{i=1}^{n-1}\alpha_i\mathbf{z}_i, \tag{A2}$$

for some $\lambda, \alpha_1, ..., \alpha_{n-1} \in \mathbb{R}$. Now, suppose that equality holds in (24). This is true if and only if

$$\left(\overline{\mathbf{W}} - \frac{\mathbf{D}\mathbf{x}_k\mathbf{x}_k^T\mathbf{D}}{\mathbf{x}_k^T\mathbf{D}\overline{\mathbf{W}}^{-1}\mathbf{D}\mathbf{x}_k}\right)\left(\mathbf{x}_{unc} - \mathbf{x}_k\right) = \mathbf{0}.$$

Using (A2), this holds if and only if $\overline{\mathbf{W}}\left(\sum_{i=1}^{n-1}\alpha_i\mathbf{z}_i\right) = \mathbf{0}$. Therefore, using (A2), equality in (24) holds if and only if $\mathbf{x}_{unc} - \mathbf{x}_k = \lambda\overline{\mathbf{W}}^{-1}\mathbf{D}\mathbf{x}_k$, for some $\lambda \in \mathbb{R}$, which is also the necessary condition in (7). The special case of $\beta_k \in \{0,1\}$ follows directly from the above also.

## 2. Proof of Proposition 2

The proof of Proposition 2 amounts to showing that $\left|d^2\phi_k/d\alpha^2\right|$ can be uniformly bounded for all $\alpha \in [0,\bar{\alpha})$ independent of choice of $\mathbf{x}_k \in \text{int}[\mathcal{L}]$. The inequality in (25) then follows from the mean-value theorem. Differentiating (21), one obtains

$$\begin{aligned}\frac{d^2\phi_k}{d\alpha^2} &= \left(\frac{d\mathbf{x}_k^s}{d\alpha}\right)^T \overline{\mathbf{W}}\mathbf{b}_k\sqrt{\frac{\ell}{\ell+\alpha^2 b}} + \frac{\alpha b}{\ell+\alpha^2 b}\mathbf{b}_k\sqrt{\frac{\ell}{\ell+\alpha^2 b}} \\ &+ \left(\frac{b}{\ell+\alpha^2 b} - 2\left(\frac{\alpha b}{\ell+\alpha^2 b}\right)^2\right)(\mathbf{x}_{unc} - \mathbf{x}_k^s(\alpha))^T\,\overline{\mathbf{W}}\mathbf{x}_k^s(\alpha) \\ &+ \frac{\alpha b}{\ell+\alpha^2 b}(\mathbf{x}_{unc} - 2\mathbf{x}_k^s(\alpha))^T\,\overline{\mathbf{W}}\frac{d\mathbf{x}_k^s}{d\alpha}. \end{aligned} \tag{A3}$$

The terms in (A3) are now uniformly bounded. First of all $\sqrt{\ell/(\ell+\alpha^2 b)} \leq 1$, since $b \geq 0$ and $\ell > 0$. Next, from (15)

$$\overline{\mathbf{W}}\mathbf{b}_k = \left[\beta_k\overline{\mathbf{W}}\left(\mathbf{I} - \frac{\mathbf{D}\mathbf{x}_k\mathbf{x}_k^T\mathbf{D}}{\|\mathbf{D}\mathbf{x}_k\|^2}\right)\overline{\mathbf{W}} + (1-\beta_k)\left(\overline{\mathbf{W}} - \frac{\mathbf{D}\mathbf{x}_k\mathbf{x}_k^T\mathbf{D}}{\mathbf{x}_k^T\mathbf{D}\overline{\mathbf{W}}^{-1}\mathbf{D}\mathbf{x}_k}\right)\right](\mathbf{x}_{unc} - \mathbf{x}_k).$$

It is well known that $\|\mathbf{I} - (\mathbf{D}\mathbf{x}_k\mathbf{x}_k^T\mathbf{D}/\|\mathbf{D}\mathbf{x}_k\|_2^2)\|_2 = 1$, and using (23) one has

$$\left\|\overline{\mathbf{W}} - \frac{\mathbf{D}\mathbf{x}_k\mathbf{x}_k^T\mathbf{D}}{\mathbf{x}_k^T\mathbf{D}\overline{\mathbf{W}}^{-1}\mathbf{D}\mathbf{x}_k}\right\|_2 \leq \|\mathbf{E}^{-1}\|_2^2.$$

By hypothesis, $\mathbf{x}_k, \mathbf{x}_k^s(\alpha) \in \mathcal{L}$, and as such they are uniformly bounded, independent of $\mathbf{x}_k \in \text{int}[\mathcal{L}]$ and $\alpha \in [0,\bar{\alpha})$. Consequently, $\overline{\mathbf{W}}\mathbf{b}_k$ is uniformly bounded independent of the choice of $\mathbf{x}_k \in \text{int}[\mathcal{L}]$. Furthermore, from (16) there exists a finite $\bar{b} > 0$ such that $0 \leq b < \bar{b}$, also independent of the choice of $\mathbf{x}_k \in \text{int}[\mathcal{L}]$. As a result, $b/(\ell+\alpha^2 b) \leq \bar{b}/\ell$. Now examine the term $(\alpha b)/(\ell+\alpha^2 b)$. Differentiating with respect to $\alpha$,

$$\frac{d}{d\alpha}\left[\frac{\alpha b}{\ell+\alpha^2 b}\right] = \frac{b(\ell-\alpha^2 b)}{(\ell+\alpha^2 b)^2},$$

shows that $(\alpha b)/(\ell+\alpha^2 b)$ is increasing for $0 \leq \alpha < \sqrt{\ell/b}$ and decreasing for $\alpha > \sqrt{\ell/b}$. Therefore, it is maximum on $\alpha \geq 0$ when $\alpha = \sqrt{\ell/b}$. Consequently, $0 \leq \alpha b/(\ell+\alpha^2 b) \leq \sqrt{b}/(2\sqrt{\ell}) \leq \sqrt{\bar{b}}/(2\sqrt{\ell})$, for all $\alpha \geq 0$, again independent of the choice of $\mathbf{x}_k \in \text{int}[\mathcal{L}]$. Using the obtained bounds on the terms in (20), there exists a finite $m_2 > 0$ such that $\|d\mathbf{x}_k^s/d\alpha\|_2 \leq m_2$, for all $\alpha \in [0,\bar{\alpha})$ independent of the choice of $\mathbf{x}_k \in \text{int}[\mathcal{L}]$. Finally, all terms in (A3) have been bounded uniformly in $\alpha \in [0,\bar{\alpha})$, independent of $\mathbf{x}_k \in \text{int}[\mathcal{L}]$. Therefore, there is a finite $M > 0$ such that $\left|d^2\phi_k/d\alpha^2\right| \leq M$, for all $\alpha \in [0, \bar{\alpha})$, independent of the choice of $\mathbf{x}_k \in \text{int}[\mathcal{L}]$.

**3. Proof of Theorem 1**

It is clear from Algorithm 1 that $\bar{J}(\mathbf{x}_{k+1}) \leq \bar{J}(\mathbf{x}_k)$ (equivalently $J(\mathbf{x}_{k+1}) \leq J(\mathbf{x}_k)$) for all $k \geq 0$. Hence, $\mathbf{x}_k \in \mathcal{J}$ for all $k \geq 0$. If for some $k \geq 0$, $\mathbf{x}_k$ is a stationary point of Problem 1, then by Proposition 1 $d\phi_k/d\alpha(0) = 0$, and consequently, $\mathbf{x}_j = \mathbf{x}_k$ for all $j \geq k$, and the result is immediate. Therefore, assume that $\mathbf{x}_k$ is not a stationary point of Problem 1 for any $k \geq 0$.

If no step length $\alpha_k \in (0, \alpha_{\max}]$ exists satisfying (27) and (28), then

$$\phi_k(\alpha_{\max}) \leq \phi_k(0) + \alpha_{\max} c_1 \frac{d\phi_k}{d\alpha}(0),$$

otherwise, using the mean value theorem, a step length satisfying (27) and (28) could be found [13, p. 35]. Using (18), (19) and (26), this can be rewritten as

$$\alpha_{\max} c_1 \left| \frac{d\phi_k}{d\alpha}(0) \right| \leq \bar{J}(\mathbf{x}_k) - \bar{J}(\mathbf{x}_{k+1}), \tag{A4}$$

since $d\phi_k/d\alpha(0) < 0$ by (22).

On the other hand, if a step length $\alpha_k \in (0, \alpha_{\max}]$ does exist satisfying (27) and (28), (27) can similarly be rewritten as

$$\alpha_k c_1 \left| \frac{d\phi_k}{d\alpha}(0) \right| \leq \bar{J}(\mathbf{x}_k) - \bar{J}(\mathbf{x}_{k+1}). \tag{A5}$$

Setting $\bar{J}_0 > \bar{J}(\mathbf{x}_0)$ in Proposition 2, one has $\mathbf{x}_k \in \mathcal{J} \subset \text{int}[\mathcal{L}]$, and the Lipschitz condition (25) gives,

$$\frac{d\phi_k}{d\alpha}(\alpha_k) \leq \frac{d\phi_k}{d\alpha}(0) + M\alpha_k.$$

Upon application of the second Wolfe condition (28),

$$c_2 \frac{d\phi_k}{d\alpha}(0) \leq \frac{d\phi_k}{d\alpha}(0) + M\alpha_k,$$

which rearranges to give a lower bound on the step length

$$\alpha_k \geq \frac{1 - c_2}{M} \left| \frac{d\phi_k}{d\alpha}(0) \right|. \tag{A6}$$

Note that it may be that $\alpha_k > \bar{\alpha}$ as defined in Proposition 2. In this case, there must exist a step length $\alpha'_k \in [0, \bar{\alpha})$ satisfying (27) and (28), and consequently also satisfying (A6). Since $\alpha_k > \alpha'_k$, (A6) still applies.

Applying (A6) to (A5) gives

$$\frac{c_1(1-c_2)}{M}\left|\frac{d\phi_k}{d\alpha}(0)\right|^2 \leq \bar{J}(\mathbf{x}_k) - \bar{J}(\mathbf{x}_{k+1}). \tag{A7}$$

From (A4) and (A7), one finds that

$$A_k \leq \bar{J}(\mathbf{x}_k) - \bar{J}(\mathbf{x}_{k+1}), \tag{A8}$$

where

$$A_k = \begin{cases} (c_1(1-c_2)/M)\left|d\phi_k/d\alpha(0)\right|^2, & \exists \alpha_k \in (0, \alpha_{\max}] \text{ satisfying (27) and (28)}, \\ \alpha_{\max} c_1 \left|d\phi_k/d\alpha(0)\right|, & \text{otherwise} \end{cases} \tag{A9}$$

Summing (A8) from $k = 0$ to $k = n$ gives $\sum_{k=0}^{n} A_k \leq \bar{J}(\mathbf{x}_0) - \bar{J}(\mathbf{x}_{n+1})$. Since $\bar{J}(\mathbf{x}_0) \geq \bar{J}(\mathbf{x}_k) \geq 0$ for all $k \geq 0$, this becomes $\sum_{k=0}^{n} A_k \leq \bar{J}(\mathbf{x}_0)$. Finally, since this must hold for all $n \geq 0$, the limit on the left must exist as $n \to \infty$ and $\sum_{k=0}^{\infty} A_k \leq \bar{J}(\mathbf{x}_0)$. Hence, $\lim_{k\to\infty} A_k = 0$, and by (A9),

$$\lim_{k\to\infty} \frac{d\phi_k}{d\alpha}(0) = 0. \tag{A10}$$

By Proposition 1, this implies that $\mathbf{x}_k$ approaches the set of stationary points of Problem 1, given by $\mathcal{C}$. Finally, noting that $\left(\mathbf{I} - \mathbf{D}\mathbf{x}_k\mathbf{x}_k^T\mathbf{D}/\|\mathbf{D}\mathbf{x}_k\|_2^2\right) = \left(\mathbf{I} - \mathbf{D}\mathbf{x}_k\mathbf{x}_k^T\mathbf{D}/\|\mathbf{D}\mathbf{x}_k\|_2^2\right)^2$ and

$$\left(\overline{\mathbf{W}} - \frac{\mathbf{D}\mathbf{x}_k\mathbf{x}_k^T\mathbf{D}}{\mathbf{x}_k^T\mathbf{D}\overline{\mathbf{W}}^{-1}\mathbf{D}\mathbf{x}_k}\right) = \left(\mathbf{I} - \frac{\overline{\mathbf{W}}^{-1}\mathbf{D}\mathbf{x}_k\mathbf{x}_k^T\mathbf{D}}{\mathbf{x}_k^T\mathbf{D}\overline{\mathbf{W}}^{-1}\mathbf{D}\mathbf{x}_k}\right)^T \overline{\mathbf{W}} \left(\mathbf{I} - \frac{\overline{\mathbf{W}}^{-1}\mathbf{D}\mathbf{x}_k\mathbf{x}_k^T\mathbf{D}}{\mathbf{x}_k^T\mathbf{D}\overline{\mathbf{W}}^{-1}\mathbf{D}\mathbf{x}_k}\right),$$

and using the fact that $0 \leq \beta_k \leq 1$, one obtains from (22)

$$\left|\frac{d\phi_k}{d\alpha}(0)\right| \geq \left\|\beta_k\left(\mathbf{I} - \frac{\mathbf{D}\mathbf{x}_k\mathbf{x}_k^T\mathbf{D}}{\|\mathbf{D}\mathbf{x}_k\|_2^2}\right)\overline{\mathbf{W}}\left(\mathbf{x}_{unc} - \mathbf{x}_k\right)\right\|_2^2 + \lambda_{\min}(\overline{\mathbf{W}})\left\|(1-\beta_k)\left(\mathbf{I} - \frac{\overline{\mathbf{W}}^{-1}\mathbf{D}\mathbf{x}_k\mathbf{x}_k^T\mathbf{D}}{\mathbf{x}_k^T\mathbf{D}\overline{\mathbf{W}}^{-1}\mathbf{D}\mathbf{x}_k}\right)(\mathbf{x}_{unc} - \mathbf{x}_k)\right\|_2^2. \tag{A11}$$

Therefore, equations (15), (22), (A10) and (A11) show that

$$\lim_{k\to\infty} \mathbf{b}_k = \mathbf{0}.$$

Finally, since $0 < \alpha_k \leq \alpha_{\max}$, (14) together with (16) and (26) give $\lim_{k\to\infty}(\mathbf{x}_{k+1} - \mathbf{x}_k) = \mathbf{0}$.

## REFERENCES


[1] Springmann, J.C., Cutler, J.W., "Optimization of Directional Sensor Orientation with Application to Sun Sensing", *Journal of Guidance, Control and Dynamics*, Vol. 37, No. 3, 2014, pp. 828–837. doi: 10.2514/1.61468

[2] O'Keefe, S.A., Schaub, H., "Sun Heading Estimation Using a Partially Undetermined Set of Coarse Sun Sensors", $2^{nd}$ IAA Conference on Dynamics and Control of Space Systems, Rome, Italy, March 24–26, 2014.

[3] Park, C., Teunissen, P.J.G., "Integer Least Squares with Quadratic Equality Constraints and Its Application to GNSS Attitude Determination Systems," *International Journal of Control, Automation, and Systems* Vol. 7, No. 4, 2009, pp. 566–576. doi: 10.1007/s12555-009-0408-0

[4] Zanetti, R., Majji, M., Bishop, R., and Mortari, D., "Norm-Constrained Kalman Filtering", *Journal of Guidance, Control and Dynamics*, Vol. 32, No. 5, 2009, pp. 1458–1465. doi: 10.2514/1.43119

[5] Wang, D., Li, M., Huang, X., Li, J., "Kalman Filtering for a Quadratic Form State Equality Constraint", *Journal of Guidance, Control and Dynamics*, Vol. 37, No. 3, 2014, pp. 951–957. doi: 10.2514/1.62890

[6] Gander, W., "Least Squares with a Quadratic Constraint," *Numer. Math.*, Vol. 36, 1981, pp. 291–307. doi: 10.1007/BF01396656

[7] Yang, C., and Blasch, E., "Kalman Filtering with Nonlinear State Constraints," *IEEE Transactions on Aerospace and Electronic Systems*, Vol. 45, No. 1, 2009, pp. 70–84. doi: 10.1109/TAES.2009.4805264

[8] Moon, T. K., and Stirling, W. C. *Mathematical Methods and Algorithms for Signal Processing* Upper Saddle River, NJ: Prentice-Hall, 2000, pp. 765–766.

[9] Elden, L., "Solving Quadratically Constrained Least Squares Problems Using a Differential-Geometric Approach," *BIT Numerical Mathematics*, Vol. 42, No. 2, 2002, pp. 323–335. doi: 10.1023/A:1021998908811

[10] Golub, G.H., von Matt, U., "Quadratically Constrained Least Squares and Quadratic Problems," *Numer. Math.*, Vol. 59, 1991, pp. 561–580. doi: 10.1007/BF01385796

[11] Chan, T.F., Olkin, J.A., Cooley, D.W., "Solving Quadratically Constrained Least Squares Using Black Box Solvers," *BIT Numerical Mathematics*, Vol. 32, No. 3, 1992, pp. 481–495. doi: 10.1007/BF02074882

[12] Crassidis, J.L., Junkins, J.L., *Optimal Estimation of Dynamic Systems*, 2nd Edition, CRC Press, Boca Raton, FL, 2012, pp. 15–16, 48–49, 67–68. ch. 5.

[13] Nocedal, J., Wright, S.J., *Numerical Optimization* Second Edition, Springer-Verlag, New York, 2006, pp. 33, 35, 60–61, 320–321,

[14] Bernstein, D., *Matrix Mathematics* Princeton University Press, New Jersey, 2009, pp. 465.

[15] Julier, S., Uhlmann, J., Durrant-Whyte, H.F., "A New Method for the Nonlinear Transformation of Means and Covariances in Filters and Estimators", *IEEE Transactions on Automatic Control*, Vol 45, No. 3, 2000, pp. 477–482. doi: 10.1109/9.847726

[16] Arasaratnam, I., Haykin, S., "Cubature Kalman Filters", *IEEE Transactions on Automatic Control*, Vol 54, No. 6, 2009, pp. 1254–1269. doi: 10.1109/TAC.2009.2019800

[17] Ainscough, T., Zanetti, R., Christian, J., Spanos, P.D., "Q-Method Extended Kalman Filter," *Journal of Guidance, Control and Dynamics*, Vol. 38, No. 4, 2015, pp. 752–760. doi: 10.2514/1.G000118

[18] de Ruiter, A.H.J., Damaren, C.J., Forbes, J.R., *Spacecraft Dynamics and Control: An Introduction*, John Wiley and Sons Ltd., 2013, ch. 25.

[19] Wertz, J.R., *Spacecraft Attitude Dynamics and Control,* Dordrecht, Holland: Kluwer Academic Publishers, 1978, App. H.

[20] Hughes, P.C., *Spacecraft Attitude Dynamics*, Dover Publications Inc., Mineola, NY, 2004, pp. 18, 236–237, 264.