Use of Constrained Nonlinear Kalman Filtering to Detect Pathological Constriction of Cerebral Arterial Blood Vessels

Federico S. Cattivelli\textsuperscript{2}, Shadnaz Asgari\textsuperscript{1}, Paul Vespa\textsuperscript{3}, Ali H. Sayed\textsuperscript{2}, Marvin Bergsneider\textsuperscript{1} and Xiao Hu\textsuperscript{1},

\textsuperscript{1}Neural Systems and Dynamics Lab, Department of Neurosurgery, University of California, \textsuperscript{2}Adaptive Systems Lab, Department of Electrical Engineering, University of California, \textsuperscript{3}Neurocritical Care Program, Department of Neurosurgery, University of California, Los Angeles, USA

1. Introduction

Aneurysmal subarachnoid hemorrhage (aSAH) is a significant health care problem because of its high morbidity and mortality rates. Survivors of initial hemorrhage are susceptible to many forms of delayed but treatable secondary injuries, among which delayed ischemic neurological deficit (DIND) caused by vasospasm is the leading cause of morbidity and mortality. It is known that between 5 and 10\% of hospitalized SAH patients die from vasospasm. What makes vasospasm interesting is that to some extent it is predictable, preventable and treatable [1].

Cerebral vasospasm is defined as the narrowing of the contrast medium column in the major cerebral arteries as evidenced in angiograms. It usually starts 3 to 5 days following bleeding showing a maximal reduction of the affected vessel lumen during Days 5 to 14 and can slowly resolve after weeks in some cases [2]. Prediction of cerebral vasospasm after aSAH is still challenging although there exist several ways of approaching it. Some patient-related factors, such as initial clinical grade of aSAH [3], size and location of aneurysm [4], age [5], and sex [6] were found predictive of cerebral vasospasm from epidemiological studies. However, these measures can be too general to be useful for individuals.

A. Detecting vasospasm

Various diagnostic neurological imaging modalities can potentially be used as indicators of vasospasm. Conventional angiography is the gold standard of confirming the narrowing of large arteries. However, it is an invasive technique, since it requires the insertion of a catheter into a peripheral artery and the addition of a dye for correct visualization. Angiography is not suitable for continuous monitoring, and can also miss small vasospastic vessels. Clinical scales, such as Fisher score [4] that quantifies total amount of subarachnoid blood on the initial CT scan, are predictive of vasospasm to some extent. However, they do not take into consideration the amount of blood removed by surgery, which has been shown to reduce the incidence of vasospasm. CT perfusion, single-photon emission computed...
tomography (SPECT) and MRI perfusion/diffusion imaging methods have the advantage of being noninvasive. However, their limitations include possible irradiation of head, cost, inconvenience of transporting ICU patients, general availability and possible delay of prediction after significant tissue ischemia is already present. More regional methods such as bedside intracerebral microdialysis have also been proposed [7], which seem appropriate for detecting focal vasospasm. However, microdialysis needs insertion of a probe to observe metabolic parameters changes in the tissue that may become ischemic [7]. In addition, an appropriate selection of which region to probe is needed. Finally, a delay of prediction may be present because of its reliance on the changed tissue metabolic patterns.

The only noninvasive and easy-to-access clinical way of predicting vasospasm, at present, is to use Transcranial Doppler (TCD) measurement of blood flow velocities in conductive cerebral arteries including middle cerebral artery (MCA), anterior cerebral artery (ACA) and intracranial carotid artery (ICA) [8], [9], [10]. Prediction of vasospasm using this conventional TCD assessment is based on absolute values of velocity (> 120 cm/s indicating medium vasospasm, > 200 cm/s indicative of severe vasospasm). These criteria are based on a very simplified view of the complex cerebral hemodynamics that cerebral flow velocity (CBFV) is inversely related to square of the vessel radius. It has been demonstrated that this relationship only partially holds when vasospasm is not severe [11]. The actual relation between CBFV and arterial radius \( r \) is as follows

\[
CBFV = \frac{CBF}{\pi r^2}
\]

where CBF is the Cerebral Blood Flow. In severe cases, a diminishing CBF is concomitant with a decreasing CBFV, which would have been taken as a recovery sign from vasospasm without knowledge of CBF. Additionally, systemic vascular effect has also been shown to affect the absolute blood flow velocity where hypertensive patients have a lower CBFV such that assessment of these patients are more prone to false negative results [12]. Another limitation of relying on the inverse relationship between vessel radii and flow velocity is that vasospasm can be detected only after certain degree of vessel radii change hence compromising its predictive power.

An alternative method that is used often in practice is the evaluation of the Lindegaard ratio [13] which is an empirical approach and does not give exact information about the actual radii of the vessels as Angiography does, and also the thresholds defined for predicting the outcome are rather ad-hoc and may change for different patients.

In this work we resort to a different approach to estimate the radii of the arteries without directly measuring them. It constitutes a model-based approach where state-estimation is applied to estimate physiological variables of interest such as arterial radii. The objective is to obtain a better estimation than that offered by the Lindegaard ratio, avoid the invasiveness of Angiography, and at the same time allow for continuous monitoring and possibly prediction of future spastic states.

**2. Methodology**

It is apparent from the above discussion that in order to increase the predictive power of TCD based vasospasm assessment, additional measurements including arterial blood
pressure (ABP) and intracranial pressure (ICP) are needed. ICP is informative for two reasons: 1) ICP directly affects cerebral perfusion pressure, whose fluctuation is one of CBF autoregulation stimuli, and autoregulation will modulate the relationship between CBFV and cerebral vasculature radii; 2) CBF changes lead to cerebral blood volume changes and thus ICP changes. Hence, ICP carries information about both CBF and cerebrospinal fluid (CSF) circulatory systems. In addition to more measurements, the dynamic changes in signals instead of absolute CBFV values shall be explored. However, what is less apparent is how to integrate all these added measurements such that dynamics can be properly characterized to provide predictive information.

Mathematical modeling of a complex system is useful in elucidating the internal factors that are causative of observed behaviors of the system. In a pure deterministic situation, such internal factors alone can determine the current state of system. The system in this study is the coupled CBF and CSF circulatory system, which is complex and distributed in nature. A mathematical model of such a system is inevitably imperfect and hence has uncertainties. Thus, when trying to assess internal states of a system from external observations, it is imperative to address the imperfectness of the model in statistical sense. This treatment then results in a model-based data fusion approach, or equivalently in an engineering term, a stochastic state estimation approach. Cerebral vasculature radii dynamics provide internal information that can potentially lead to a more straightforward and efficient way of predicting vasospasm offering considerable advantage than predicting vasospasm based on any single measurement of the system.

Three components are needed to construct such a state estimator for vasospasm prediction as described in our previous work [14]:

- A mathematical model of cerebral hemodynamics (Sec. III)
- A model-training (or parameter estimation) approach (Sec. IV)
- A nonlinear state-estimation approach (Sec. V)

The first component is a mathematical model of the integrated cerebral blood and CSF circulatory systems. Among many published models, those proposed in a series work of Ursino [15], [16], [17], [18], [19], [20], [21], [22] are appropriate ones because all the key known physiological factors are included in these models that regulate CBF and their interactions with the CSF circulation. However, one issue still remains regarding the complexity of the model if one chooses to model the vasospasm directly. Trade-off is usually justified between the complexity and the accuracy of the model because a state estimator is easier to be built on a simpler model as there would be fewer parameters to be estimated and the numerical stability of the estimator would be better. We approached this problem by a simulation study where signals simulated using a complete model including a direct modeling of vasospasm are used to test a state estimator based on a simplified model without components of vasospasm. This will be illustrated in Section VI.

The second component is the parametrization of the model using individual patient’s measurement. This is essentially a nonlinear optimization problem. We will briefly describe in Section IV our solution that utilizes a combined global random search of parameter space and a local gradient based search to fine tune the parameters found by the global search. Particularly, we chose the differential evolution (DE) algorithm in the present work [23]. This algorithm has been used successfully as a global search technique for parameterizing mathematical models encountered in various fields [24], [25], [26].

www.intechopen.com
Once an individualized model is obtained, the third step is to simulate the model and adjust its estimated state variables based on the error between the simulated and the measured output of the model. The adjustment of state variable is necessary because any inaccuracy of the model, any physiological changes of parameters with time, or the accumulated error of numerical integrator can drive the simulated results erroneous. The celebrated Kalman filter (KF) [27] is a better solution because errors sensed in the simulated output are then used to correct the state estimation at every measurement moment. The KF is an optimal state estimator for linear Gaussian systems. Even though it is well known that such optimality is lost for a nonlinear dynamic system, almost all of the recently proposed nonlinear state estimators still follow [28], [29], [30], [31] the KF’s general schema to achieve a suboptimal solution for such systems although they differ in the particular form of propagating the statistics of state variables. In essence, one could chose any of the above nonlinear estimators for the vasospasm detection problem. However, all these filters have usually derived the Kalman gain in an unconstrained fashion, meaning that no domain knowledge has been used to define a feasible range of solutions for states and model parameters. In an intracranial pressure dynamic model, such physiological constraints usually exist for state variables as well as for model parameters. Therefore, a regularization of the state estimation process can be achieved by incorporating constraints in the derivation of the Kalman gain. A quadratic programming technique can be used to solve the resultant constraint optimization problem, which is applicable to any nonlinear state estimators.

In summary, we propose a model-based approach to integrate ABP, ICP, and CBFV signals so that a hidden variable useful for characterizing the narrowing process of cerebral arteries can be obtained for detecting this process before any clinical symptoms appear. This approach is comprised of a physiology-driven mathematical model of CBF and CSF circulatory systems, a parameter estimation module that provides appropriate values for unknown model parameters, and a state updating algorithm based on KF-like nonlinear estimators (see Fig. 1). In the following, brief technical descriptions of modeling, parametrization, and state estimation are provided. A simulation study will be introduced to illustrate the feasibility of using a simplified model to drive the state estimation process. The effectiveness of the whole system will be illustrated using data from patients who developed vasospasm.

All the variables mentioned in this work correspond to time domain signals, sampled at 1Hz. The mathematical models used in this work have inputs, outputs, state variables and parameters. The input in this case is Arterial Blood Pressure (ABP), and the outputs are Intracranial Pressure (ICP) and Cerebral Blood Flow Velocity (CBFV). We assume measurements of all inputs and outputs are available. The models have several parameters which are in general unknown. An example of a parameter is the nominal value of a vessel resistance. The states typically represent some physiological variables which may not be measured directly, such as arterial radii of the vessels, compartment compliances, etc, and therefore need to be estimated.

Note that the parameter estimation and state estimation could be combined into one step, by posing the problem as a joint parameter-state estimation. Note however that this requires knowing the dynamics of the model parameters a-priori. Another technique that makes the same assumption is dual estimation. None of these topics are discussed in this work, and the reader is referred to [32] for more information.
3. Mathematical models of cerebral hemodynamics

We will introduce a complete dynamic model of CSF circulation, CBF circulation, and cerebral vasospasm. This model will be simulated with realistic parameters to provide ground truth data for validating the proposed state estimation approach. We will then present a systematic way of simplifying this model at different levels of complexity.

The proposed methodology for the estimation of arterial radii is based on continuous time measurements of CBFV, ABP and ICP. This methodology relies heavily on mathematical models that relate these quantities, together with the desired arterial radii. For our purpose, a good mathematical model should provide good correlation with observed quantities, and at the same time have low complexity to allow fast training and state estimation, and avoid possible instability. In general, these two characteristics will contradict each other, i.e., a less complex model will be less able to capture the interrelations between all the variables.

Another limitation of the approach is that even if we have a good model that closely matches the observed variables, it is virtually impossible to obtain continuous measurements of the actual arterial radii to compare them with the estimates. Hence, in this work we propose a simulation-based approach as follows: we develop a mathematical model of cerebral hemodynamics that is more general than previous models, and takes into account mechanisms such as Autoregulation and vasospasm. We will denote this model as Model 1. Then, we will use Model 1 to generate artificial data for different values of spasm severity. Next, we will develop a second model, denoted as Model 2, to estimate the arterial radii from Model 1 based on its outputs. As mentioned before, we want Model 2 to be simple, in order to reduce the complexity of the parameter and state estimation. This simulation-based approach will give us good insight into how capable simple models are of predicting states from more complex ones, and is the first step towards the application of the state estimation on actual patient data.
The mathematical models derived in this work are based on the models proposed by Ursino et al. These models were first introduced in [16], [17], [18]. Our work is based on the model of [16]. One inconvenience of the model in [16] is that it does not model vasospasm, which makes it inappropriate for the generation of data at different levels of spasm. Vasospasm was modeled in the work by Lodi and Ursino [19], but unfortunately several simplifications were introduced to the original Ursino model, such as a much simpler Autoregulation mechanism, and collapsing of the small and large arterial sections into one single section. Hence, we combined the two aforementioned models into one more general model that takes into account vasospasm, has a detailed Autoregulation mechanism, and has four sections: namely those corresponding to the large arteries (MCA, ACA, PCA), followed by the large pial arteries, small pial arteries and capillaries, and finally the venous compartment. We refer to this model as Model 1, and present it in the form of an electrical circuit in Fig. 2.

![Model 1](image)

Fig. 2. Model 1, which is a combination of models published in [16] and [19].

Next we introduced several simplifications to Model 1, namely collapsing small and large pial arterial sections into one, a simpler Autoregulation mechanism, and assuming $P_v = P_{ic}$. We also added one capacitance at the large arteries to obtain a state variable that allowed us to obtain the desired MCA radius. We refer to this model as Model 2, and present it in the form of an electrical circuit in Fig. 3.

![Model 2](image)

Fig. 3. Model 2, which is a simplified version of Model 1.
Use of Constrained Nonlinear Kalman Filtering to Detect Pathological Constriction of Cerebral Arterial Blood Vessels

The details of models 1 and 2 can be found in [33] and are omitted from this work due to space considerations. Slight modifications were introduced to the models in [33] and will appear in future publications.

4. Parameter estimation

Fig. 1a shows the Model Training scenario. A model such as the ones described in Section III is used to generate artificial outputs (ICP and CBFV), and the measurements of these outputs are subtracted to generate an error signal. An optimization block is used to select the set of parameters that minimizes some cost function that depends on the error. For instance, in our case we use the cost function

$$J(\theta) = \sum_{l=1}^{L} \sum_{n=1}^{N} w_l(n) [y_l(n) - \hat{y}_l(n, \theta)]^2$$

(1)

where \(N\) is the total number of measurements, \(L\) is the total number of outputs, \(\theta\) is the unknown parameter vector, \(y_l(n)\) is the \(n\)th measurement of output \(l\), \(\hat{y}_l(n, \theta)\) is the \(n\)th output \(l\) generated by the model using parameter \(\theta\), and \(w_l(n)\) is some weighting function. In our case, we use the weighting function that weights every variable inversely proportional to the square of signal sample \(y_l(n)\).

The models considered in this work are nonlinear, and hence (1) will in general be a nonconvex function of \(\theta\). As such, algorithms based on gradient descent are not guaranteed to converge to a global optimum. Hence, the optimization is done in two steps as proposed in [14]. First, a global search is performed using a genetic algorithm known as Differential Evolution (DE) [23], which has low complexity and good convergence. After the global search, a local search is performed using a standard gradient descent algorithm through the MATLAB Optimization Toolbox. Further details of this approach can be found in [14].

5. Non-linear state estimation

A. Kalman Filter-like state estimation for nonlinear models

Extensive treatments of the Kalman filter can be found in many standard text books [34], [35], [36], [37]. The extension of the Kalman Filter to nonlinear systems is required for detecting cerebrovascular radii changes. Consider a mathematical model of the system as follows

$$\mathbf{x}_{n+1} = F(\mathbf{x}_n, \mathbf{u}_n, \mathbf{w}_n)$$

(2)

and its measurement function as

$$\mathbf{y}_n = H(\mathbf{x}_n, \mathbf{u}_n, \mathbf{v}_n)$$

(3)

where \(\mathbf{x}_n, \mathbf{u}_n, \) and \(\mathbf{w}_n\) are the state variables, input, and state noise, respectively at time instant \(n\). \(\mathbf{y}_n\) and \(\mathbf{v}_n\) are the model output and observation noise, respectively. For the continuous-time systems studied here (see Section 3), \(\mathbf{x}_{n+1}\) is obtained by the numerical integration procedure with \(\mathbf{x}_n, \mathbf{u}_n\) and \(\mathbf{w}_n\) as inputs.
The optimal estimate of the state variables $x_n$ in the least mean squares sense, given the observation of $y_i, i = 1, ..., n$, is the conditional expectation $E(x_n | y_i, i = 1, ..., n)$. The most significant contribution of the KF is that it realizes a recursive procedure to obtain this conditional expectation exactly for a linear system with jointly Gaussian initial states and noise variables. However, this optimality is not retained in general for nonlinear systems. Instead, several extensions of the linear Kalman Filter can be used to obtain a suboptimal solution for nonlinear systems. A typical approach to solve the problem is to use the Extended Kalman Filter (EKF), which has the disadvantage of requiring the Jacobian matrix of the system, its calculation being error prone. Derivative-free state estimation approaches in non-linear systems have also been proposed, for example, the Unscented Kalman Filter [38] and the DD1 and DD2 filters [31], which have been shown to provide better performance than the EKF. We give a brief introduction of the general paradigm of the Kalman filter in this section so that the links between the nonlinear and linear Kalman filters can be clearly observed.

In the subsequent mathematical developments, the notation $\hat{x}_{i|j}$ is used to denote an estimate of $x_i$ based on the measurements up to discrete time $j$, and the operator $^*$ denotes conjugate transposition.

Let $\hat{x}_{n|n-1}$ denote the prior estimation of $x_n$, namely

$$\hat{x}_{n|n-1} = E[x_n | y_0, \ldots, y_{n-1}]$$

where $E[\cdot]$ is the expectation operator, and define the error quantity $\tilde{x}_{n|n-1} = x_n - \hat{x}_{n|n-1}$ with its associated covariance matrix as

$$P_{\tilde{x}_{n|n-1}} = E((x_n - \tilde{x}_{n|n-1})(x_n - \tilde{x}_{n|n-1})^*)$$

Similarly, let $\tilde{y}_{n|n-1}$ denote the prior estimation of $y_n$ and define the innovation sequence

$$e_n = y_n - \tilde{y}_{n|n-1}$$

with covariance

$$R_{e,n} = E e_n e_n^*$$

At time $n$, a new measurement $y_n$ is collected to derive a better estimation of $x_n$. For nonlinear filters, the assumption is that the random variables $x_n$ and $y_n$ are jointly Gaussian. When two random variables $x$ and $y$ are jointly Gaussian, it is well known that the optimal estimator of $x$ given $y$ is affine [36], [39]. Thus, in the case of a Gaussian system, it can be shown that the estimate of $\hat{x}_{n|n}$ can be obtained from $\hat{x}_{n|n-1}$ and the difference between the measured $y_n$ and $\tilde{y}_{n|n-1}$ such that

$$\hat{x}_{n|n} = \hat{x}_{n|n-1} + K_n (y_n - \tilde{y}_{n|n-1})$$

where $K_n$ is termed Kalman gain, which, for both linear and nonlinear systems, can be optimally calculated as

$$K_n = E[\hat{x}_{n|n-1} e_n] R_{e,n}^{-1}$$

where $E[\hat{x}_{n|n-1} e_n]$ is the cross covariance between $\hat{x}_{n|n-1}$ and $e_n$. Moreover, the state error covariance can be updated from
Use of Constrained Nonlinear Kalman Filtering to Detect Pathological Constriction of Cerebral Arterial Blood Vessels

151

Equation (4), used to obtain $\hat{x}_{n|n}$ from $\hat{x}_{n|n-1}$, is usually called measurement update since the upgrade of the prior estimate to a better posterior estimate is achieved with the arrival of a new measurement. The optimal choices for the remaining quantities are:

$$\hat{x}_{n+1|n} = E[F(x_n, u_n, w_n)|y_0, \ldots, y_n]$$ (6)

which requires the calculation of the conditional probability of $x_n$ given the measurements up to time $n$. This step is usually called time-update. The prediction of the measurement can be calculated in a similar fashion as

$$\hat{y}_{n|n-1} = E[H(x_n, u_n, v_n)|y_0, \ldots, y_{n-1}]$$ (7)

Having obtained $\hat{x}_{n+1|n}$, $\hat{y}_{n+1|n}$ and their covariances, the next run of measurement-updates can be carried out. The whole iteration procedure then continues. The above equations allow us to recursively compute the state estimates $\hat{x}_{n|n}$. Note that we have not assumed linearity of the model, and our only assumption was that the states and measurements are jointly Gaussian.

The time update step in the general Kalman filter paradigm is essentially the propagation of the expectation and the covariances of random variables through functions. A difficulty arises in general when calculating (6) and (7) since the value of $x_n$ is unknown. Different nonlinear Kalman filters address this propagation problem in different ways while the measurement update is conducted in the same fashion. Nonlinear Kalman-like estimators such as the Extended Kalman Filter (EKF) use the following approximations

The Unscented Kalman filter [38], [32] uses the so-called sigma points, $\chi_{0,i} = \hat{x}_{i|n}$, $\chi_{n,i} = \hat{x}_{i|n} + \sqrt{(L + \lambda)P_{i|n}}$, $n = 1, \ldots, L$ and $\chi_{n,i} = \hat{x}_{i|n} - \sqrt{(L + \lambda)P_{i|n}}$, $n = L + 1, \ldots, 2L$, where $\lambda = \alpha^2(L + \kappa) - L$, $\alpha$ and $\kappa$ are two constant parameters. Then the sigma points are propagated through the nonlinear function $f$, and the resulting points are aggregated to obtain an approximation for the mean of the output of the function, as follows:

$$\chi_{n,i+1|i} = f(\chi_{n,i+1|i}, u_k)$$

$$\hat{x}_{i+1|i} \approx \sum_{n=0}^{2L} W_n \chi_{n,i+1|i}$$

In the next two sections, we will briefly describe a different approach based on the DD1 filter [31].

**B. DD1 filter**

DD1 and DD2 filters were introduced by Norgaard [31], and approximate the nonlinear transformations using a multidimensional extension of Stirling’s interpolation formula. As in Unscented filters, the DD1 and DD2 filters do not require calculation or evaluation of the derivatives of the nonlinear functions. Also, they have been shown to outperform the EKF in several situations.
The basic principle behind DD1 and DD2 filters is that a nonlinear function \( f(x) \) can be approximated as

\[
f(x) \approx f(\bar{x}) + f'_D(\bar{x})(x - \bar{x}) + \frac{1}{2} f''_D(\bar{x})(x - \bar{x})^2 \ldots
\]

where

\[
f'_D(x) = \frac{f(x+h) - f(x-h)}{2h} \quad \text{and} \quad f''_D(x) = \frac{f(x+h) + f(x-h) - 2f(x)}{h^2}
\]  

(8)

Given a random variable \( x \), with mean \( E(x) = \bar{x} \) and covariance \( \text{Cov}(x) = E(x-\bar{x})(x-\bar{x})^* = P_x \), and an arbitrary function \( f(.) \), we would like to determine \( E(y) = \bar{y} \) and \( \text{Cov}(y) = P_y \), where \( y = f(x) \). In other words, we are interested in propagating the mean and covariance of a random variable through a possibly nonlinear function \( f \). When \( f \) is linear, the answer is given by the Kalman filter.

Define a random variable \( z = S^{-1}_x x \), where \( S_x \) is a Cholesky factor of \( P_x \), i.e., \( P_x = S_x S^*_x \).

Then we have that \( P_{zz} = I \). Moreover, define a function \( \tilde{f}(z) = f(S^{-1}_x x) = f(x) \). To first order, we can approximate:

\[
y = \tilde{f}(z) = \tilde{f}(\bar{z}) + \tilde{f}'(\bar{z})\Delta z
\]

(9)

where \( \Delta z \triangleq z - E(z) \) is a zero-mean random variable. Thus, taking expectation of (9), to first order we obtain:

\[
\bar{y} \approx f(\bar{x})
\]

(10)

In order to propagate the covariance matrix, the same first order approximation leads us to

\[
P_y = E[(y - \bar{y})(y - \bar{y})^*] \approx E[\tilde{f}'_D(\bar{z})\Delta z(\tilde{f}'_D(\bar{z})\Delta z)^*]
\]

Noting that \( E(\Delta z) = I \), and using (8) we obtain

\[
P_y \approx \frac{1}{4h^2} [f(\bar{x} + hS_x) - f(\bar{x} - hS_x)][f(\bar{x} + hS_x) - f(\bar{x} - hS_x)]^*
\]

In the general case where \( x \) has size \( n \), [31] shows that \( P_y \) and the cross covariance \( P_{xy} \) can be obtained from

\[
P_y \approx \frac{1}{4h^2} \sum_{p=1}^{n} [f(\bar{x} + h s_{x,p}) - f(\bar{x} - h s_{x,p})][f(\bar{x} + h s_{x,p}) - f(\bar{x} - h s_{x,p})]^*
\]

(11)

\[
P_{xy} \approx \frac{1}{2h} \sum_{p=1}^{n} s_{x,p}[f(\bar{x} + h s_{x,p}) - f(\bar{x} - h s_{x,p})]^*
\]

(12)

where \( s_{x,p} \) is the \( p^{th} \) column of \( S_x \).

The DD1 filter uses the same approximation as the EKF for the a-priori update, namely
\[ \hat{x}_{n+1|n} = F(\hat{x}_{n|n}, u_n, \bar{w}_n) \]

and the output estimate is given by

\[ \hat{y}_{n|n-1} = H(\hat{x}_{n|n-1}, u_n, \bar{v}_n) \]

The Kalman gain is computed from

\[ K_n = P_{xe,n}R_{e,n}^{-1} \]

where \( R_{e,n} \) and the cross covariance \( P_{xe,n} \) are obtained from (11) and (12), respectively, by propagating \((\hat{x}_{n|n-1}, u_n, v_n)\) through the function \( H \). For these calculations, an estimate of \( P_{x|n-1} \) is required, which is obtained from (11) by propagating \((\hat{x}_{n-1|n-1}, u_{n-1}, w_{n-1})\) through the function \( F \). The measurement update is the same as in the linear case, namely,

\[ \hat{x}_{n|n} = \hat{x}_{n|n-1} + K_n(y_n - \hat{y}_{n|n-1}) \]

\[ P_{\hat{x}n|n} = P_{\hat{x}n|n-1} - K_nR_{e,n}K_n^T \]

C. Constrained Kalman Filter

Physical and physiological boundaries on some state variables and model parameters exist in the present model. However, it can be noted that the equation to update state variables can result in posterior estimates well beyond those boundaries. Hence, adding constraints to the original problem of optimizing \( K_n \) is necessary. It is shown that this can be formulated as a quadratic programming problem for which efficient algorithms exist. Quadratic programming solves the following problem

\[ \min_{\theta \in \mathbb{R}^n} J(\theta) = \frac{1}{2} \theta^T H \theta + \omega^T \theta \]

\[ A_i \theta = b_i, \quad i = 1, \cdots, n_e \]

\[ A_i \theta \leq b_i, \quad i = n_e + 1, \cdots, n_p \]

where \( H, \omega, A, \) and \( b \) are known quantities. \( A_i \) is the \( i \)th row of constraint matrix \( A \) and \( b_i \) is the \( i \)th element of \( b \). In the above formulation, there are \( n_e \) equality and \( n_p - n_e \) inequality constraints. To formulate the calculation of \( K_n \) as a QP problem, it should be realized that \( K_n \) is an optimal solution of linear least mean square problem, i.e.,

\[ \min_{K_n} Tr E[\hat{x}_n - \hat{x}_{n|n-1} - K_n (y_n - \hat{y}_{n|n-1})][\hat{x}_n - \hat{x}_{n|n-1} - K_n (y_n - \hat{y}_{n|n-1})]^T \]

where \( Tr[ ] \) is the trace of a matrix. The above equation can be further simplified since all the variables involved are real. With definitions of \( \hat{x} = x_n - \hat{x}_{n|n-1} \) and \( \hat{y} = y_n - \hat{y}_{n|n-1} \), it follows

\[ \min_{K_n} Tr \left[ E\hat{x}\hat{x}^T + K_n E\hat{y}\hat{y}^T K_n^T - 2K_n E\hat{y}\hat{x}^T \right] \]

Denoting the \( i \)th row of \( K_n \) as \( K_{n,i} \) and the dimension of state variable as \( d \), the above problem can be decomposed into \( d \) independent subproblems as
\[
\min_{K_{n,i}} E\hat{x}_i^2 + K_{n,i}E\hat{y}^T K_{n,i}^T - 2K_{n,i}E\hat{y}\hat{x}_i^T
\]

which is equivalently

\[
\min_{K_{n,i}} \frac{1}{2} K_{n,i}E\hat{y}^T K_{n,i}^T + (-K_{n,i})E\hat{y}\hat{x}_i^T
\]

hence \( \theta \) in Eq 14 is \( -K_{n,i}^T \). Now suppose that the \( x_n \) is constrained by \([l^{(i)}, u^{(i)}]\), which leads to

\[
l^{(i)} \leq \hat{x}_{n|n-1,i} + K_{n,i}(y_n - \hat{y}_{n|n-1}) \leq u^{(i)},
\]

simple manipulations lead to the standard form of inequity constraints as

\[
K_{n,i}(y_n - \hat{y}_{n|n-1}) \leq u^{(i)} - \hat{x}_{n|n-1,i}
\]

\[
K_{n,i}(\hat{y}_{n|n-1} - y_n) \leq \hat{x}_{n|n-1,i} - l^{(i)}.
\]

This clearly indicates that \( K_{n,i} \) can be calculated by solving a QP problem instead of the original unconstrained least mean square solution.

6. Numerical simulation

Our goal is to use a mathematical model of cerebral hemodynamics, combined with a state-estimation approach which, using signals produced by a human patient, will allow us to estimate the hidden states of the patient. In Section 3 we discussed two models of cerebral hemodynamics. Model 1 is a more general version of the model presented in [16], where it was shown that the results obtained through this model are in accordance with expected human responses. This model has ten state variables and several unknown parameters. Thus, if we were to use Model 1 for our state-estimation approach, the technique would become computationally complex and error prone, making Model 1 a poor candidate for our state estimation approach.

Model 2 is a simpler model, with only four state variables, namely \( P_{la}, P_{pa}, P_{ic} \) and \( C_{pa} \). Thus this model is more amenable to our estimation approach and will be more robust to model mismatch.

It is, however, interesting to analyze how accurate is Model 2 in modeling the human cerebral hemodynamics. In other words, will Model 2 be able to detect changes in arterial radius as well as Model 1 would? To answer this question, we propose the following approach. We assume that Model 1 is the “true” model, and use it to generate artificial ICP and CBFV signals at different spastic levels. Subsequently, we apply our state-estimation approach using Model 2, on the signals generated by Model 1, and use the resulting state estimates to obtain estimates of arterial radius.

Figure 4 shows the estimation results (from [33]). The solid blue curve shows the actual value of the radius introduced artificially in Model 1, and the red dashed curve shows the radius estimated by Model 2 through measurement of the ICP and CBFV signals produced by Model 1. We observe that Model 2 is able to accurately track the changes in radius. Thus Model 2 is a good candidate for our state estimation approach.
In the following section we will show the estimation results obtained using real patient data instead of artificial data generated by a model. We will use Model 2 as our state-estimation model.

7. Results with real patient data

We now present our state estimation results using real patient data. We show results on two patients who suffered Subarachnoid Hemorrhage (SAH), and were admitted to the Neurocritical care unit at the UCLA Medical Center in 2007. Both patients have continuous recordings of ICP, ABP and CBFV measured at the right MCA, on different days after the initial SAH. Typically, two recording sessions were obtained each day, one in the morning and one in the afternoon. The TCD monitoring and data collection were approved by the UCLA IRB with signed consent form from patients’ next-of-kin.

The length of each recording varies, but typically includes between 10 minutes to 1 hour of usable data. The first patient, which we will refer to as “Patient A”, had recordings on days 1, 3, 4, 5, 7, and 8 after the initial SAH. Days 1, 3, 4 and 8 have two sessions each. We will denote each recording session by the letter of the patient followed by the day after SAH, and then the session number on that day. Thus, the recordings of Patient A are labelled as A1-1, A1-2, A3-1, A3-2, A4-1, A4-2, A5-1, A7-1, A8-1 and A8-2. We will refer to the second patient as “Patient B”, which has recordings B4-2, B5-1, B5-2, B6-1, B8-1, B10-1. This patient had an additional recording on day 4 after SAH, which was excluded since it did not include enough clean data for processing, due to either movement artifacts or noisy use of the TCD probe.

For each recording, we processed the data as follows. For each session, we extracted 300 seconds of clean data, including continuous ICP, CBFV at the right MCA and ABP. The 300 seconds were extracted by observing that the signals were clean enough, that there were no transient artifacts due to placement of the TCD probe, and that no significant movement artifacts were present. After we extracted the signals, we downsampled the signals to 1Hz sampling rate, previously using a lowpass filter to avoid aliasing. Note that this downsampling process keeps only average information of the signals, and the pulsatility information is destroyed.
After we obtained the downsampled signals, we applied our parameter estimation approach on the 300 seconds obtained from the first session available for that patient. That is, we used session A1-1 to estimate the parameters for patient A, and session B4-2 to estimate the parameters of patient B. Then we kept those same initial parameters constant for all the remaining sessions. The rationale behind this choice is that in the earliest recording, the patient is more likely not to have developed vasospasm. Typically vasospasm takes 3 to 4 days to develop. Thus, for patient A, it is very likely that the recording on day 1 will not include the effects of vasospasm. Then we will obtain a set of parameters that is representative of the “normal” state, without spasm. Then, a subsequent recording with a spastic state should be detected since it would deviate from the “normal” conditions. Unfortunately, for patient B, the earliest recording available is on day 4, indicating that the spasm may be present already (though we will argue later that this is not the case). Thus, if we train in a state that already includes the spasm, we may not be able to detect the subsequent changes. In summary, it is important to train the model as soon as possible after the initial SAH, to allow the system to learn the model parameters in normal conditions.

After learning the model parameters on the first available recording, we proceeded to estimate the states of the model using two different approaches: a DD1 filter, and a DD1 filter with constraints on the states, denoted by QCKF. The details of the constraints will be presented in future publications.

Figure 5 shows the estimated radius for every session of patient A, both for the DD1 and QCKF filters. We can observe that according to our predictions, this patient had an evolving vasospasm starting on day 1, session 2, after the initial SAH. This spasm get accentuated on day 4, where it reaches less than half of its initial value. Then there is a slight recovery up to day 7, though the spasm is still considered severe. On day 8, session 2, the radius of the artery recovers to normal values, indicating the possibility of an angioplasty (surgery used to return arterial radius to normal).

![Fig. 5. Estimated radii for all sessions of patient A, for DD1 and QCKF filters.](image)

Figure 6 shows the measured ICP and CBFV against the signals estimated through our state estimation approach, both for DD1 and QCKF. We can observe that for both methods, the observed and estimated ICP have good agreement. This is not the case for CBFV, where
some discrepancies are observed. For instance, for session A7-1, DD1 has a significant
disagreement with the actual signals. Note, however, how the QCKF approach is able to
reduce this difference, by constraining the states to be within allowed values. On session A8-
2, both filters have a significant difference with the true values, which is due to the fact that
the CBFV is significantly increased from the normal values, thus questioning whether the
estimates obtained in Fig. 5 for session A8-2 are accurate or not. Note that the two filters also
disagree on the radius estimates of session A8-1. The output estimates provided by the
QCKF are closer to the actual measurements on this session, thus we could conclude that the
QCKF is providing better estimates of the radius, and that the angioplasty was performed
before session A8-1.

Note also that the changes in ICP and CBFV do not seem to follow the trend of the
vasospasm, which shows how powerful the proposed technique is. The mathematical
models employed capture several complex inter-relationships between ICP, ABP and CBFV
that are not evident at simple sight.

![Graph showing actual and estimated outputs (ICP, CBFV) for patient A, using DD1 filter (top two plots) and QCKF filter (bottom two plots).]

Fig. 6. Actual and estimated outputs (ICP, CBFV) for patient A, using DD1 filter (top two plots) and QCKF filter (bottom two plots).
The estimation results obtained for patient B are shown in Fig. 7, for the QCKF filter. The results obtained through DD1 filtering are similar. It can be observed that the patient did not suffer from significant changes in arterial radius. This indicates that either the patient did not suffer from vasospasm after SAH, or that the patient already had vasospasm on day 4 where the training was made, and the spasm did not improve in subsequent days. Angiographic evidence shows that the patient did have a mild vasospasm, supporting the second hypothesis.

![Fig. 7. Estimated radius for patient B, using QCKF filters.](image1)

Fig. 8 shows the measured and estimated outputs for Patient B, where a good match is observed for all recordings. Note again how big changes in CBFV and ICP do not influence the arterial radius estimates considerably, as is expected from this patient.

![Fig. 8. Actual and estimated outputs (ICP, CBFV) for patient B, using the QCKF filter.](image2)
8. Discussion and conclusions

Estimation of lumped cerebral arterial radius is important for healthcare monitoring in ICU patients, specially for detecting the presence of vasospasm following Subarachnoid Hemorrhage. A technique for continuous monitoring based on available measurements without introducing any additional invasive technique is very attractive, and would allow detection of vasospasm earlier and more accurately than other methods such as angiography.

Our proposed estimation approach uses a combination of parameter estimation and state estimation techniques, and relies heavily on mathematical models of cerebral hemodynamics. We presented two models based on previous work by Ursino et. al, and showed through simulation how a simpler Model 2, with only four state-variables could predict changes in arterial radius from the signals generated through Model 1. We showed how to estimate the parameters of Model 1 through a non-linear least-squares technique, and how we trained the model on the first available recording of every patient. Then we applied our state-estimation approach using DD1 filtering, and a DD1 filter with constraints (QCKF). We showed that our approach detected the presence of vasospasm for Patient A, and observed a radius evolution that matches the expected results. The QCKF produced smaller errors in the output estimation, indicating that constraining the states to be within reasonable limits may improve the accuracy of the estimation. We also estimated the arterial radius for Patient B, which had a constant mild vasospasm throughout the recordings.

In essence, we have shown the potential of Kalman-like state-estimators for nonlinear models. This application could potentially save lives by predicting post-SAH vasospasm before other techniques would.

9. References


Use of Constrained Nonlinear Kalman Filtering to Detect Pathological Constriction of Cerebral Arterial Blood Vessels


The aim of this book is to provide an overview of recent developments in Kalman filter theory and their applications in engineering and scientific fields. The book is divided into 24 chapters and organized in five blocks corresponding to recent advances in Kalman filtering theory, applications in medical and biological sciences, tracking and positioning systems, electrical engineering and, finally, industrial processes and communication networks.

**How to reference**

In order to correctly reference this scholarly work, feel free to copy and paste the following:
