2011 American Control Conference
on O'Farrell Street, San Francisco, CA, USA
June 29 - July 01, 2011

Bochner Integrable Solutions to Riccati Partial Differential Equations and Optimal Sensor Placement

J. A. Burns * and C. N. Rautenberg †

* Interdisciplinary Center for Applied Mathematics, Virginia Tech, Blacksburg, VA jaburns@vt.edu

† Interdisciplinary Center for Applied Mathematics, Virginia Tech, Blacksburg, VA rautenbe@vt.edu

Abstract— **In this paper we provide sufficient conditions to ensure that solutions to the time varying Riccati partial differential equations are Bochner integrable with range in the space of trace class operators. The fact that Bochner integrals can be uniformly approximated by simple functions provides a basis for obtaining bounds on integration errors. These bounds can then be used for rigorous numerical analysis and to ensure the convergence of algorithms used to compute approximate solutions. We demonstrate how this result can be employed to develop convergent computational methods for a sensor placement problem based on optimal filtering. Theoretical results are presented and numerical examples are given to illustrate the ideas.**

I. Introduction and Problem Formulation

Let $\Omega \subset \mathbb{R}^n$, $n = 1, 2, 3$, be an open bounded domain with boundary $\partial\Omega$ of Lipschitz class (for example the open unit cube in $\mathbb{R}^n$ has Lipschitz class boundary). Consider a advection-diffusion process in the region Ω with boundary $\partial\Omega$ described by the partial differential equation with disturbance given by

$$\frac{\partial}{\partial t}T(t,\vec{x}) = \epsilon^2 \Delta T(t,\vec{x}) + [\kappa(\vec{x}) \cdot \nabla T(t,\vec{x})] \quad (1)$$

$$+ \sum_{k=1}^{m} g_k(\vec{x})\eta_k(t), \quad (2)$$

with boundary and initial conditions

$$T(t,\vec{x}) \mid_{\partial\Omega} = 0, \quad T(0,\vec{x}) = T_0(\vec{x}) \in L^2(\Omega).$$

Here, the functions $g_k(\cdot)$ are given and $\eta_k(\cdot)$ represents a time varying disturbance so that

$$g(t,x) = \sum_{k=1}^{m} g_k(\vec{x})\eta_k(t) \quad (3)$$

is a spatially distributed disturbance.

We assume that there is a single "mobile sensor platform" that produces a local spatial average of the state $T(t,\vec{x})$. The extension to several sensors is straightforward. Thus, we assume that $h(\cdot)$ is a given weighting function and a sensor is moving along the path described by $\vec{\gamma}(t) = [x(t), y(t), z(t)]^T \in \Omega$ which produces an spatially averaged signal on a ball of radius δ about the trajectory $\vec{\gamma}(t)$. Consequently, the measured output has the form

$$y(t) = \iiint\limits_{B_\delta(\vec{\gamma}(t))\cap\Omega} h(\vec{x})T(t,\vec{x})d\vec{x} + v(t), \quad (4)$$

where $v(t)$ is the sensor noise and $B_\delta(\vec{\gamma}(t))$ is the open ball about the curve $\vec{\gamma}(t)$ of radius δ.

For the mobile sensor defined by $\vec{\gamma}(t)$, we define the output map $C(t) : L^2(\Omega) \longrightarrow \mathbb{R}^1$ by

$$C(t)\varphi(\cdot) = \iiint\limits_{B_\delta(\vec{\gamma}(t))} h(\vec{x})\varphi(\vec{x})d\vec{x} \quad (5)$$

and hence the measured output defined by (4) has the abstract form

$$y(t) = C(t)T(t,\cdot) + Ev(t), \quad (6)$$

where in this one sensor case $E = 1$.

The standard formulation of the abstract (infinite dimensional) model for the convection-diffusion system (2) with output (6) leads to the a distributed parameter system in the Hilbert space $\mathscr{H} = L^2(\Omega)$ given by

$$\dot{z}(t) = Az(t) + G\vec{\eta}(t) \in \mathscr{H}, \quad (7)$$

with output

$$y(t) = C(t)z(t) + Ev(t), \quad (8)$$

where the state of the distributed parameter system is $z(t)(\cdot) = T(t,\cdot) \in \mathscr{H} = L^2(\Omega)$ and A is the usual convection-diffusion operator (see [1], [2]).

We focus here on a version of the optimal sensor management problem first formulated by Bensoussan and Curtain (see [4], [5], [8], [9]). In particular, the problem is formulated as an optimal estimation problem based on the observation that the variance equation for the Kalman filter satisfies an infinite dimensional Riccati (partial) differential equation of the form

$$\dot{\Sigma}(t) = A\Sigma(t) + \Sigma(t)A^* \quad (9)$$
$$+ GG^* - \Sigma(t)C^*(t)C(t)\Sigma(t),$$

with initial data

$$\Sigma(t_0) = \Sigma_0. \quad (10)$$

The optimal sensor management problem becomes a distributed parameter optimal control problem with the "state" $\Sigma(\cdot)$ defined by the Riccati system (9)-(10) and the cost functional is defined in terms of the trace of $\Sigma(\cdot)$. In particular, we fix a time interval $0 \le t \le t_f$ and assume $Q(\cdot) = Q(\cdot)^* \in L^\infty([0,t_f];\mathscr{L}(H))$ is a given self-adjoint weighting operator. The optimal mobile sensor problem can be stated as the following optimal control problem (see [6]):

978-1-4577-0079-8/11/$26.00 ©2011 AACC

Problem A: Find $C^{opt}(\cdot)$ of the form (5) so that

$$J\big(C(\cdot)\big) = \int_{t_0}^{t_f} \mathrm{Tr}(Q(t)\Sigma(t))dt \tag{11}$$

is minimized, where $\Sigma(\cdot)$ **is a solution of the system** (9)-(10) **and** $C(\cdot)$ **is of the form** (5).

Remark A. Note that this problem could have been stated in terms of finding the optimal trajectory $\vec{\gamma}^{opt}(t) = [x(t), y(t), z(t)]^T$. In practice, the trajectories are defined (limited) by the dynamics of the platform. Also, the way one defines what is meant by a solution to the Riccati differential equation (9)-(10) can dramatically effect the choice of a numerical approximation, the convergence and convergence rates. We turn to the integral equation form of (9)-(10).

II. The Riccati Integral Equation

In much of the literature, solutions of (9)-(10) are defined in terms of mild forms of the integral equation

$$\begin{aligned}\Sigma(t) = S^*(t)\Sigma_0 S(t) + \int_0^t S^*(t-s) \\ \times (G(s)G^*(s) - \Sigma(s)(C^*(s)C(s))\Sigma(s)) \\ \times S(t-s)ds.\end{aligned} \tag{12}$$

Under the assumptions considered here, we have shown that (9)-(10) can be interpreted as a Bochner integral equation with operator-valued integrand. Since Bochner integrable functions can be uniformly approximated by simple functions (e.g. step functions), one can obtain (uniform) bounds on numerical integration and thus control the integration errors to ensure convergence of the integral in (11) and for the corresponding approximating solutions. It is shown in [16] that controlling these tolerances is the key to obtaining correct numerical approximations of the cost function (11).

Let $\mathscr{H}$ be a separable complex Hilbert space. The spaces $\mathscr{I}_1$ and $\mathscr{I}_2$ denote the space of trace class and Hilbert-Schmidt operators on $\mathscr{H}$, respectively. The following result is a special case of Proposition 11 in [16].

Theorem 1. *Let $I = [0, t_f]$ and assume*

(i) *$S(t)$ is a C_0-semigroup over $\mathscr{H}$;*
(ii) *$\Sigma_0 \in \mathscr{I}_1$;*
(iii) *$F(\cdot) = G(\cdot)G^*(\cdot) \in L^1(I; \mathscr{I}_1)$;*
(iv) *$D(\cdot) = C(\cdot)^*C(\cdot) \in L^\infty(I; \mathscr{L}(H))$;*

hold. If $\Sigma(\cdot) \in L^2(I; \mathscr{I}_2)$, then for all $t \in I$ the mapping

$$s \mapsto S(t-s)(F(s) - \Sigma(s)D(s)\Sigma(s))S^*(t-s)$$

is Bochner integrable and $\Upsilon(\Sigma)(\cdot)$ defined by

$$\begin{aligned}\Upsilon(\Sigma(\cdot))(t) = S(t)\Sigma_0 S^*(t) + \int_0^t S(t-s) \\ \times (F(s) - \Sigma(s)D(s)\Sigma(s)) \\ \times S^*(t-s)ds,\end{aligned}$$

is a well defined function $\Upsilon : L^2(I; \mathscr{I}_2) \mapsto \mathscr{C}(I; \mathscr{I}_1)$. Moreover, since

$$\Upsilon\Big(L^2(I; \mathscr{I}_2)\Big) \subset \mathscr{C}(I; \mathscr{I}_1),$$

it follows that $\mathscr{C}(I; \mathscr{I}_1)$ is an Υ-invariant subspace of $L^2(I; \mathscr{I}_2)$.

If in addition, $D(\cdot)$ satisfies the stronger condition

(iv') *$D(\cdot) \in L^\infty(I; \mathscr{I}_1)$,*

and $\Sigma(\cdot) \in L^2(I; \mathscr{L}(\mathscr{H}))$, then $\Upsilon(\Sigma)(\cdot) \in \mathscr{C}(I; \mathscr{I}_p)$. In this case,

$$\Upsilon\Big(L^2(I; \mathscr{L}(\mathscr{H}))\Big) \subset \mathscr{C}(I; \mathscr{I}_1),$$

and $\mathscr{C}(I; \mathscr{I}_1)$ is a Υ-invariant subspace of $L^2(I; \mathscr{L}(\mathscr{H}))$.

Remark B. The benefit of using a Bochner integral is clearly demonstrated in Chapter 5 of the thesis [16] where it is shown that numerically incorrect answers can be obtained unless one can set specific tolerances on these Bochner integrals. These issues are discussed in detail in the thesis [16] and will appear in a future longer paper.

In the next section we show that under reasonable assumptions, Galerkin and spectral type numerical methods produce convergence in the trace norm and these schemes provide a practical algorithm for solving the optimal sensor management problem.

III. Approximation

One of the main issues that needs to be addressed in order to develop practical numerical schemes is the approximating the operator $C(\cdot)$. Here, $C(\cdot)$ is assumed to vary continuously with respect to the Hilbert-Schmidt norm. This implies that $C^*(\cdot)C(\cdot)$ varies continuously with respect to the trace norm, and this provides the foundation that allows us to work with trace class solutions of the Riccati equation. We will make use of known results on existence and approximations developed by [10], [11] and [12]. These references deal with convergence in the space of Hilbert-Schmidt operators $\mathscr{I}_2$. We will extend these results to convergence in the space $\mathscr{I}_1$ of trace class operators.

A. Assumptions and Basic Results

Let, for each $n \in N$ let P_n be the projection from the Hilbert space $\mathscr{H}$ onto a finite dimensional Hilbert space $\mathcal{V}_n$ such that $\mathcal{V}_n \subset \mathscr{H}$ and $\mathcal{V}_n \subset \mathscr{D}(A)$, where the sequence $P_n^* P_n$ converges strongly to the identity and $[\mathcal{N}(P_n)]^\perp \subset \mathscr{D}(A)$ for each $n \in N$. Since$P_n^* P_n$ converges strongly to the identity and $\mathcal{V}_n \subset \mathscr{H}$, the sequence $\|P_n\|$ is uniformly bounded.

Also, $A_n = P_n A P_n^*$ is a bounded linear operator on $\mathcal{V}_n \subset \mathscr{H}$ and is an infinitesimal generator of the uniformly continuous semigroup $T_n(t) = e^{A_n t}$ on $\mathscr{H}$. Let $T(t)$ denote the semigroup with infinitesimal generator A and assume the following conditions hold.

H1) There exist $M \geq 1$ and $\omega \in \mathbb{R}$ such that $\|T_n(t)\| \leq Me^{\omega t}$.
H2) There is a dense subset $D \subset \mathscr{H}$ such that $D \subset \mathscr{D}(A)$ and such that if $x \in D$, then $A_n x \to Ax$ as

$n \to \infty$. Moreover, there is a complex number λ_0, with $\mathbf{Re}\,\lambda_0 > \omega$ such that $\overline{(\lambda_0 - A)D} = \mathscr{H}$.

If **H1** and **H2** are satisfied, then the Trotter-Kato Theorem (see [2] and [15]) implies that as as $n \to \infty$

$$\|T(t)x - T_n(t)x\| \to 0, \tag{13}$$

for each $x \in \mathscr{H}$ and the convergence is uniform on compact time intervals. When P_n is an orthogonal projection, ($P_n^* = P_n$) and $P_n \to I$ strongly the following result holds.

Proposition 1. *Let $\{P_n\}_{n=1}^{\infty}$ be a sequence of orthogonal projectors over a complex separable Hilbert space $\mathscr{H}$ that converge strongly to the identity, $0 \le \Sigma_0 \in \mathscr{I}_1$, $F(\cdot) \in L^1(I;\mathscr{I}_1)$ and $D(\cdot) \in \mathscr{C}(I;\mathscr{I}_1)$. If $I = [0, t_f]$, then*

i. *$P_n\Sigma_0 P_n \in \mathscr{I}_1$ and $\|\Sigma_0 - P_n\Sigma_0P_n\|_1 \to 0$ as $n \to \infty$.*

ii. *The map $t \mapsto P_nF(t)P_n$ belongs to $L^1(I;\mathscr{I}_1)$ and*
$$\int_I \|(F - P_nFP_n)(s)\|_1 \,\mathrm{d}s \to 0,$$
as $n \to \infty$.

iii. *The map $t \mapsto P_nD(t)P_n$ belong to $\mathscr{C}(I;\mathscr{I}_1)$ and*
$$\sup_{t\in I} \|(D - P_nDP_n)(t)\|_1 \to 0,$$
as $n \to \infty$.

Proof:

i. Let $\Sigma_0 \ge 0$ be of rank one, so $\Sigma_0 x = \langle \varphi, x\rangle \varphi$ and $P_n\Sigma_0P_nx = \langle P_n\varphi, x\rangle P_n\varphi$ since $P_n^* = P_n$, for all $x \in \mathscr{H}$. Define $\varphi_n = P_n\varphi$. It follows that $(\Sigma_0 - P_n\Sigma_0P_n)x = \langle\varphi, x\rangle\varphi - \langle\varphi_n, x\rangle\varphi_n = \langle\varphi - \varphi_n, x\rangle\varphi + \langle\varphi_n, x\rangle(\varphi - \varphi_n)$ and
$$\begin{aligned}|\mathrm{Tr}(D(\Sigma_0 - P_n\Sigma_0P_n)| \le& \\ &\sum_{k=1}^{\infty} |< \varphi - \varphi_n, \varphi_k >| \\ &\times |< \varphi_k, D\varphi >| \\ &+ \sum_{k=1}^{\infty} |< \varphi_n, \varphi_k >| \\ &\times |< \varphi_k, D(\varphi - \varphi_n) >| \\ &\le \|\varphi - \varphi_n\|\,\|D\varphi\| \\ &+ \|\varphi_n\|\,\|D(\varphi - \varphi_n)\| \\ &\le \|D\varphi\|\,\|\varphi - \varphi_n\|\,\|\varphi\| \\ &+ \|D\varphi\|\,\|\varphi_n\|\,\|\varphi - \varphi_n\|.\end{aligned}$$
Therefore,
$$\begin{aligned}\|\Sigma_0 - P_n\Sigma_0P_n\|_1 =& \\ &\sup_D \frac{|\mathrm{Tr}\,(D(\Sigma_0 - P_n\Sigma_0P_n))\,|}{\|D\|} \\ &\le \|\varphi - \varphi_n\|\|\varphi\| \\ &\quad + \|\varphi_n\|\|\varphi - \varphi_n\|,\end{aligned}$$
where the sup is over all finite rank D. This implies that $\|\Sigma_0 - P_n\Sigma_0P_n\|_1 \to 0$ since $\varphi_n \to \varphi$. If Σ_0 is of finite rank, the same result follows easily. If $0 \le \Sigma_0 \in \mathscr{I}_1$, then there is a sequence of finite rank operators $\{\Sigma_0^n\}_{n=1}^{\infty}$ such that $\|\Sigma_0 - \Sigma_0^n\|_1 \to 0$ as $n \to \infty$. This yields
$$\begin{aligned}\|\Sigma_0 - P_n\Sigma_0P_n\|_1 \le& \|\Sigma_0 - \Sigma_0^m\|_1 \\ &+ \|\Sigma_0^m - P_n\Sigma_0^mP_n\|_1 \\ &+ \|P_n(\Sigma_0 - \Sigma_0^m)P_n\|_1 \\ \le& 2\,\|\Sigma_0 - \Sigma_0^m\|_1 \\ &+ \|\Sigma_0^m - P_n\Sigma_0^mP_1\|_1,\end{aligned}$$
and hence $\limsup_{n\to\infty} \|\Sigma_0 - P_n\Sigma_0P_1\|_1 \le 2\,\|\Sigma_0 - \Sigma_0^m\|_1$. Consequently, it now follows that $\|\Sigma_0 - P_n\Sigma_0P_n\|_1 \to 0$ as $n \to \infty$ since $\|\Sigma_0 - \Sigma_0^m\|_1 \to 0$ as $m \to \infty$.

ii. Since P_n is not time dependent, it is easy to show that $P_nF(\cdot)P_n \in L^1(I;\mathscr{I}_1)$. If $F(t)$ is the step function given by $F(t) = \sum_{k=1}^{q} f_k\chi_{I_k}(t)$, then
$$\begin{aligned}&\int_I \|(F - P_nFP_n)(t)\|_1 \mathrm{d}t \\ &\qquad \le \sum_{k=1}^{q} \|f_k - P_nf_kP_n\|_1 m(I_k),\end{aligned}$$
and from the previous result we have that $\int_I \|(F - P_nFP_n)(t)\|_1 \to 0$ as $n \to \infty$. Since, step functions are dense in $L^1(I;\mathscr{I}_1)$ the result will follow for any $F(\cdot) \in L^1(I;\mathscr{I}_1)$.

iii. Since I is compact, then step functions are dense in $\mathscr{C}(I;\mathscr{I}_1)$ and this implies that the result holds in all $\mathscr{C}(I;\mathscr{I}_1)$ and this completes the proof.

B. The Convection-Diffusion Operator case

As above, let $\Omega \subset \mathbb{R}^n$ be an open bounded domain with boundary $\partial\Omega$ of Lipschitz class and consider the differential operator of order 2

$$A(x, D) = -\epsilon^2\Delta + \sum_{|\alpha|\le 1} a_\alpha(x)D^\alpha,$$

with $\epsilon > 0$ and where Δ is the Laplacian operator on Ω and the functions $x \mapsto a_\alpha(x)$ are smooth complex valued functions on $\overline{\Omega}$. Since $\epsilon^2 > 0$ then $A(x, D)$ is strongly elliptic of order 2 (see [15]) and the operator $-A = -A(x, D)$ with domain $\mathscr{D}(-A(x, D)) = H^2(\Omega) \cap H_0^1(\Omega)$ generates a C_0-semigroup $T(t) = e^{-At}$ on $\mathscr{H} = L^2(\Omega)$. The unique solution to

$$\begin{aligned}\frac{\partial u(t,x)}{\partial t} + A(x, D)u(t, x) = 0, &\quad \text{for } t > 0 \text{ and } x \in \Omega \\ u(t, x) = 0, &\quad \text{for } t \ge 0 \text{ and } x \in \partial\Omega \\ u(0, x) = u_0(x), &\quad \text{for } u_0(\cdot) \in L^2(\Omega),\end{aligned}$$

is given by

$$u(t, x) = (T(t)u_0)(x).$$

It is a well know fact that the Laplacian defined as

$$\Delta : H^2(\Omega) \cap H_0^1(\Omega) \mapsto L^2(\Omega),$$

has eigenvalues $\{\lambda_k\}_{k=1}^{\infty}$ that can be arranged in decreasing order $0 \geq \lambda_1 \geq \lambda_2 \geq \cdots$ such $\lambda_k \to -\infty$ as $k \to \infty$ and the eigenspaces are finite-dimensional. We can choose the eigenfunctionss $\{\phi_k(\cdot)\}_{k=1}^{\infty}$ to be an orthonormal basis of $L^2(\Omega)$. Moreover, $\{\phi_k(\cdot)\}$ are of class $C^{\infty}(\Omega)$.

Define the subspaces

$$\mathcal{V}_n = \operatorname{span}\{\phi_1, \phi_2, \ldots, \phi_n\},$$

and let P_n be the orthogonal projector from $L^2(\Omega)$ to $\mathcal{V}_n$. Clearly, $\mathcal{V}_n \in \mathscr{D}(-A)$ and $P_n^* P_n = P_n^2 = P_n \to I$ strongly as $n \to \infty$ since

$$\|(I - P_n)\psi\|^2 = \sum_{k=n+1}^{\infty} |\langle \phi_k, \psi \rangle|^2 \to 0, \qquad \text{as } n \to \infty,$$

and

$$\begin{aligned}\big(\mathcal{N}(P_n)\big)^{\perp} &= \big(\overline{\operatorname{span}\{\phi_{n+1}, \phi_{n+2}, \ldots\}}\big)^{\perp} \\ &= \operatorname{span}\{\phi_1, \phi_2, \ldots, \phi_n\} \\ &= \mathcal{V}_n \\ &\subset \mathscr{D}(-A).\end{aligned}$$

We show that these Galerkin approximations satisfy the conditions required by the Trotter-Kato Theorem. First, it follows from Garding's inequality that there is a $\hat{\lambda}_0 \geq 0$ such that $-A_{\hat{\lambda}_0} = -(A + \hat{\lambda}_0 I)$ is the infinitesimal generator of a C_0-semigroup of contractions in $L^2(\Omega)$ (see [15]). In particular, $-A_{\hat{\lambda}_0} \in \mathbf{G}(1,0)$. Moreover, $-A = -A_{\hat{\lambda}_0} + \hat{\lambda}_0 I$ and since $\hat{\lambda}_0 I$ is bounded with $\|\hat{\lambda}_0 I\| = \hat{\lambda}_0$, it follows that $-A \in \mathbf{G}(1, \hat{\lambda}_0)$. Since $\mathcal{V}_n \in \mathscr{D}(A)$, then it follows that $A_n = P_n A P_n$ satisfies $-A_n \in \mathbf{G}(1, \hat{\lambda}_0)$ (see [2] for a detailed proof). This observation implies that hypothesis **H1** is satisfied, since

$$\|T_n(t)\| \leq e^{\hat{\lambda}_0 t} \text{ for all } n \in \mathbb{N} \text{ and all } t \geq 0,$$

where $T_n(t)$ is the uniformly continuous semigroup generated by $-A_n$. In addition $\|T(t)\| \leq e^{\hat{\lambda}_0 t}$ for all $t \geq 0$ where $T(t)$ is the semigroup generated by $-A$.

We now show that Hypothesis **H2** is also satisfied. Let D be the space spanned by finite linear combinations of the $\{\phi_k\}_{k=1}^{\infty}$ given by

$$D = \operatorname{span}\{\phi_1, \phi_2, \ldots\}.$$

Since $\{\phi_k\}_{k=1}^{\infty}$ is an orthonormal basis of $L^2(\Omega)$, it follows that D is dense in $L^2(\Omega)$. If $x \in D$, then $x = \sum_{k=1}^{N} \langle \phi_k, x \rangle \phi_k$ for some $N < \infty$, Thus, if $n \geq N$, then

$$\begin{aligned}\|Ax - A_n x\| &\leq \sum_{k=1}^{N} |\langle \phi_k, x \rangle| \|A\phi_k - A_n \phi_k\| \\ &\leq \sum_{k=1}^{N} |\langle \phi_k, x \rangle| \|A\phi_k - P_n A \phi_k\| \to 0\end{aligned}$$

as $n \to \infty$ since $P_n \to I$ strongly as $n \to \infty$ and $N < \infty$.

Finally, we must show that there is a complex number λ_0 with $\mathbf{Re}\,\lambda_0 > \hat{\lambda}_0$ such that $\overline{(\lambda_0 + A)D} = L^2(\Omega)$. This can be done exactly as in Pazy's book (see [15]) where Pazy discusses Parabolic Equations and will not repeated here. Consequently, we have shown that **H1** and **H2** are satisfied when we used the Galerkin scheme defined by the approximations generated by $P_n(-A)P_n$ for $n = 1, 2, \ldots$. Hence it follows that

$$\|T(t)x - T_n(t)x\| \to 0,$$

as $n \to \infty$, for each $x \in L^2(\Omega)$.

Since $A(x, D) = -\epsilon^2 \Delta + \sum_{|\alpha| \leq 1} a_{\alpha}(x) D^{\alpha}$ its formal adjoint $A^*(x, D)$ is defined by (see [15])

$$A^*(x, D)u = -\epsilon^2 \Delta u + \sum_{|\alpha| \leq 1} D^{\alpha}\big(\overline{a_{\alpha}(x)}u\big),$$

and it is also strongly elliptic of order 2. Since the infinitesimal generator $-A$ is defined by $Ax = A(x, D)x$ for each $x \in H_0^1(\Omega) \cap H^2(\Omega)$, its adjoint is given by $A^* x = A^*(x, D)x$ for each $x \in H_0^1(\Omega) \cap H^2(\Omega)$ (for a proof see Pazy's book [15]). Therefore, it is straight forward to show that for each $x \in L^2(\Omega)$

$$\|T^*(t)x - T_n^*(t)x\| \to 0,$$

as $n \to \infty$, where $T^*(t)$ is the C_0-semigroup generated by $-A^*$ and $T_n^*(t)$ are the uniformly continuous semigroup generated by $P_n(-A^*)P_n$. This established conditions **H1** and **H2** are valid for the Galerkin approximations above.

We note that a a general abstract approximation framework was first developed by Banks and Kunisch in [1] for a much wider class of parabolic systems using a similar approach. Finally, we have the following theorem for trace norm convergence of the approximate solutions to the Riccati equation (9)-(10).

Theorem 2 (BASIC APPROXIMATION THEOREM). *Let X be a complex separable Hilbert space and let Y be a complex finite dimensional Hilbert space. Let $T(t)$ be the C_0-semigroup on $\mathscr{H} = L^2(\Omega)$ generated by the strongly elliptic operator $-A$ and let $T_n(t)$ be the sequence generated by $-A_n = P_n(-A)P_n$. Suppose also that*

i. $0 \leq \Sigma_0 \in \mathscr{I}_1(\mathscr{H})$.
ii. $G(\cdot) \in L^2([0, t_f]; \mathscr{I}_2(X, \mathscr{H}))$.
iii. $C(\cdot) \in \mathscr{C}([0, t_f]; \mathscr{L}(\mathscr{H}, Y))$.

Then , $\Sigma(\cdot) \in \mathscr{C}([0, t_f]; \mathscr{I}_1(\mathscr{H}))$, is the unique solution of Bochner integral equation

$$\begin{aligned}\Sigma(t) = {} & T^*(t)\Sigma_0 T(t) \\ & + \int_0^t T^*(t-s)(GG^* \\ & \qquad - \Sigma(s)(C^*C)(s)\Sigma(s))T(t-s)\mathrm{d}s,\end{aligned}$$

and the sequence of solutions $\Sigma_n(\cdot) \in \mathscr{C}([0, t_f]; \mathscr{I}_1(\mathscr{H}))$ of

$$\begin{aligned}\Sigma_n(t) = {} & T_n^*(t)\big(P_n \Sigma_0 P_n\big)T_n(t) \\ & + \int_0^t T_n^*(t-s)\Big(\big(P_n G G^* P_n\big) \\ & \qquad - \Sigma_n\big(P_n C^* C P_n\big)\Sigma_n\Big)(s) T_n(t-s)\mathrm{d}s,\end{aligned}$$

satisfies

$$\sup_{t \in [0, t_f]} \|\Sigma(t) - \Sigma_n(t)\|_1 \to 0, \tag{14}$$

as $n \to \infty$.

A detailed proof of this result is too long for this short paper, but can be found in the thesis [16]. Finally, we have to address the problem of convergence of the integrals $\int_0^{t_f} \mathrm{Tr}\left(Q\Sigma_n\right)(t)dt$ to $\int_0^{t_f} \mathrm{Tr}\left(Q\Sigma\right)(t)dt$. This is a consequence of the following Corollary.

Corollary III.1. *Assume the hypothesis of the previous Theorem 2 hold and suppose that $Q(\cdot) \in L^\infty([0,t_f];\mathscr{L}(H))$. Then the sequence $\Sigma_n(\cdot)$ and $\Sigma(\cdot)$ defined above satisfies*

$$\int_0^{t_f} \mathrm{Tr}\left(Q\Sigma_n\right)(t)\,\mathrm{d}t \to \int_0^{t_f} \mathrm{Tr}\left(Q\Sigma\right)(t)\,\mathrm{d}t.$$

Proof: The proof follows immediately from the inequality

$$|\int Q(\Sigma-\Sigma_n)(t)dt| \leq \sup_{t\in[0,t_f]} \|(\Sigma-\Sigma_n)(t)\|_1 \int_0^{t_f} \|Q(t)\|dt.$$

IV. Numerical Results for the Galerkin Approximation Scheme: Stationary Case

We present some typical numerical results for a 2D problem on the unit square. In particular, we consider

$$\frac{\partial T}{\partial t} = \epsilon^2 \Delta T + \left(a_x \frac{\partial T}{\partial x} + a_y \frac{\partial T}{\partial y}\right) + g(x,y)\eta(t),$$

with a fixed sensor is at position (x_0, y_0). The output is given by

$$y(t) = \int_\Omega h(x-x_0, y-y_0) T(t,x,y)\,\mathrm{d}x\,\mathrm{d}y + \nu(t),$$

where the kernel $h(x,y)$ is given by

$$h(x,y) = e^{-20(x^2+y^2)} \tag{15}$$

and $\epsilon^2 = 0.01$. The functional to minimize is defined by

$$J(x_0,y_0) = \int_0^1 \mathrm{Tr}\left(\Sigma_{(x_0,y_0)}(t)\right)\mathrm{d}t,$$

where $\Sigma_{(x_0,y_0)}$ is the solution of the Riccati equation and the output map is determined by the sensor in position (x_0,y_0).

Integration of the finite dimensional approximating Riccati equation was computed by an implicit Euler's method and the integral of the trace was computed using trapezoidal integration (see [3]). In all cases presented here, the Galerkin method converged with 16 basis elements. Also, to check the numerical results we compared the Galerkin scheme to the standard finite element method. In both cases, the algorithms produced the same optimal sensor location. However, the finite element method required many more basis elements before convergence was observed.

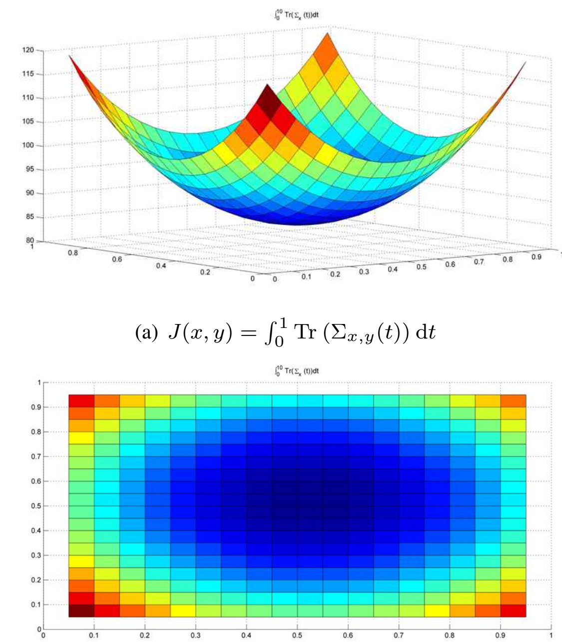
The figures below contain plots and contour plots of the cost function

$$J(x,y) = \int_0^1 \mathrm{Tr}\left(\Sigma_{x,y}(t)\right)\mathrm{d}t.$$

for $(x,y) \in \Omega$.

Uniform Noise and Zero Convective Term

In this run we set $g(x,y) = g_1(x,y) = 50$ and $a_x = a_y = 0$. In Figure 1 we plot the cost function $J(x,y) = \int_0^1 \mathrm{Tr}\left(\Sigma_{x,y}(t)\right)\mathrm{d}t$. As expected, we observe that the minimum is obtained by placing the sensor in the center of the unit square.



(a) $J(x,y) = \int_0^1 \mathrm{Tr}\left(\Sigma_{x,y}(t)\right)\mathrm{d}t$

(b) $J(x,y) = \int_0^1 \mathrm{Tr}\left(\Sigma_{x,y}(t)\right)\mathrm{d}t$ Top view

Fig. 1. $J(x,y)$ for noise $g_1(x,y) = 50$ and $a_x = a_y = 0$.

Non-uniform Noise and Zero Convective Term

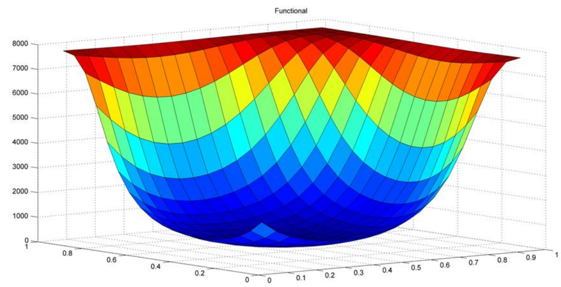
Here we set the noise term to be $g(x,y) = g_2(x,y) = 10 + 40exp[-5\left((x-0.1)^2 + (y-0.1)^2\right)]$ and again consider the zero convection $a_x = a_y = 0$ case. Here we observe that the minimum is now located at a point that lies between the center of the square and the point with the highest concentration of noise located at $(0.1, 0.1)$. Figure 2 contains the plots for this case.

Uniform Noise and Non-zero Convective Term

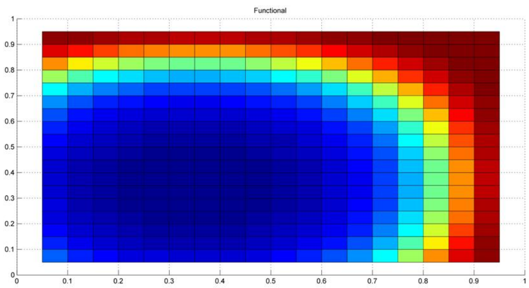
Again we assume a uniform noise and set $g(x,y,z) = g_3(x,y) = 50$. However, we allow for convection in the x-direction so that $a_x = 10$ and $a_y = 0$. Observe that this means that the field is convecting from right to left. This is the most interesting case in that the optimal location now moves to the right from the center point to an "up stream" point closer to the $y = 1$ boundary. Figure 3 contains the plots for this case.

V. Conclusions

We have provided a proof that Galerkin type approximations yield trace norm convergence of the approximating Riccati operators. As noted in [13], this type of strong convergence is required if one hopes to use numerical approximations for the optimal sensor placement problem

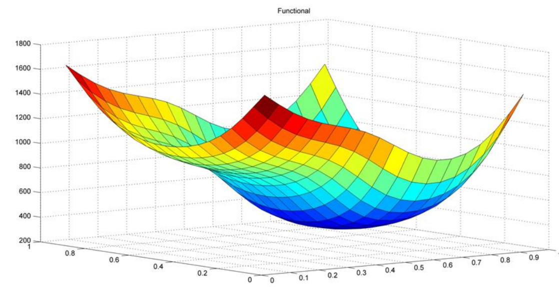


(a) $J(x,y) = \int_0^1 \mathrm{Tr}\,(\Sigma_{x,y}(t))\,\mathrm{d}t$

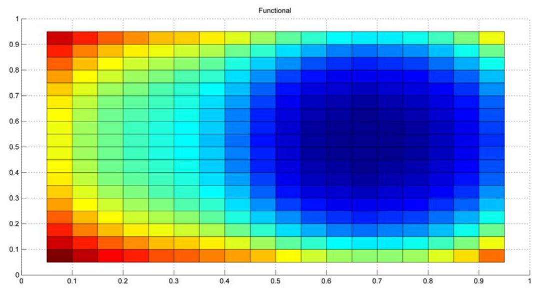


(b) $J(x,y) = \int_0^1 \mathrm{Tr}\,(\Sigma_{x,y}(t))\,\mathrm{d}t$ Top view

Fig. 2. $J(x,y)$ for noise $g(x,y) = g_2(x,y)$ and $a_x = a_y = 0$.



(a) $J(x,y) = \int_0^1 \mathrm{Tr}\,(\Sigma_{x,y}(t))\,\mathrm{d}t$



(b) $J(x,y) = \int_0^1 \mathrm{Tr}\,(\Sigma_{x,y}(t))\,\mathrm{d}t$ Top view

Fig. 3. $J(x,y)$ for noise $g_3(x,y) = 50$, $a_x = 10$ and $a_y = 0$.

based on the optimal filtering formulation. A standard finite element scheme produced the same optimal sensor location as the Galerkin algorithm, but required more elements. Note that the standard finite element scheme fails to satisfy the conditions in the approximation theorem above and hence we do not have a complete trace norm convergence theory for the finite element method. However, since the finite element scheme produced the same optimal location we conjecture there may be other conditions to ensure convergence sufficient for the sensor placement problem. This remains an open question.

The third numerical example seems to suggest that the optimal sensor location is consistent with the "intuitive answer" that the sensor should be placed up stream to deal with "lag" due to the convection. This behavior occurs because the goal here is to estimate the entire state $T(t,\vec{x})$ over the entire spatial domain Ω. Previous results (see [6] and [7]) indicate that this intuitive solution no longer remains valid if one restricts the spatial domain to a subset of Ω.

Acknowledgements This research was supported in part by Air Force Office of Scientific Research grants FA9550-07-1-0273 and FA9550-10-1-0201 and by the Environmental Security Technology Certification Program (ESTCP) under a subcontract from United Technologies.

REFERENCES

[1] H. T. Banks and K. Kunisch *The Linear Regulator Problem for Parabolic Systems*, SIAM J. Controls and Optimization, Vol 22, No. 5, pp. 684-698, 1984.

[2] H. T. Banks and K. Kunisch *Estimation Techniques for Distributed Parameter Systems*, Birkhäuser, 1989.

[3] P. Benner and H. Mena, BDF Methods for Large-Scale Differential Riccati Equations, TU Chemnitz, preprint.

[4] A. Bensoussan, *Optimization of Sensors' Location in a Distributed Filtering Problem*, Lecture Notes in Mathematics no. 294, Springer, New York, pp. 62-84.

[5] A. Bensoussan, *Filtrage Optimal des Systemes Lineaires*, Dunod, Paris, 1975.

[6] J. A. Burns, E. M. Cliff, and C. Rautenberg, *A distributed parameter control approach to optimal filtering and smoothing with mobile sensor networks*, in 17th Mediterranean Conference on Control and Automation, Thessaloniki, Greece, June 2009, 181–186.

[7] J. Borggaard, J. Burns, A. Surana and L. Zietsman, *Control, Estimation and Optimization of Energy Efficient Buildings*, in Proceedings of the 2009 American Control Conference, St. Louis, MO, June 2009, 837–841.

[8] R. F. Curtain, *A Survey of Infinite-Dimensional Filtering*, SIAM Review, 17 (1975), 395-411.

[9] R. F. Curtain, *Infinite-Dimensional Filtering*, SIAM J. Control , 13 (1975), 89-104.

[10] G. Da Prato, *Linear Quadratic Control Theory for Infinite Dimensional Systems*, Lecture notes, Summer School on Mathematical Control Theory, Trieste, 2001.

[11] A. De Santis, A. Germani and L. Jetto, Approximation of the Algebraic Riccati Equation in the Hilbert Space of Hilbert-Schmidt Operators, *SIAM J. Control Opt.*, 31 (1993), 847–874.

[12] A. Germani, L. Jetto and M. Piccioni, Galerkin Approximation for Optimal Linear Filtering of Infinite-Dimensional Linear Systems, *SIAM J. Control and Opt.*, 26 (1988), 1287–1305.

[13] K. Morris, LQ-Optimal Actuator Location and Norm Convergence of Ricatti Operators, In *Proc. of the 47th IEEE Control and Decision Conference*, Dec.2008, 209–214.

[14] S.Omato, S. Koide, and T Soeda, Optimal Sensor Location Problem for a Linear Distributed Parameter System, *IEEE Trans. on Auto. Control* 23 (1978), 665-673.

[15] A. Pazy, *Semigroups of Linear Operators and Applications to Partial Differential Equations*, Springer-Verlag, 1983.

[16] C. N. Rautenberg, *A distributed parameter control approach to optimal filtering and estimation with mobile sensor networks*, Ph.D. Thesis, Department of Mathematics, Virginia Tech, Blacksburg, VA, March, 2010.