PARAMETER ESTIMATION AND WAVEFORM DESIGN FOR COGNITIVE RADAR BY MINIMAL FREE-ENERGY PRINCIPLE

Anish Turlapaty and Yuanwei Jin

Department of Engineering and Aviation Sciences
University of Maryland Eastern Shore
Princess Anne, MD 21853

ABSTRACT

In this paper we develop a new framework for Bayesian parameter estimation using adaptive waveforms by the minimal free energy (FE) principle in the context of cognitive radar. Unlike conventional approaches, the new method utilizes the minimal FE principle as a unifying criterion for optimal estimator design and waveform design. The FE principle seeks to approximate the true density of the unknown parameters in response to sequential measurement data. In the case of a single unknown parameter we show that the estimators based on the FE principle and the conventional Bayesian estimator are identical. Moreover, the waveform design based on the FE principle results in similar water-filling solution as the traditional mutual information method.

***Index Terms*—** Cognitive Radar, Free-Energy Principle, Adaptive Waveform, Machine Learning

1. INTRODUCTION

Cognitive radar [1] is considered as an intelligent active sensing system that utilizes adaptive radar waveforms (e.g. [2, 3]) and machine learning techniques to achieve improved performance for radar tasks such as target recognition [4, 5], sensor scheduling [6] and scene analysis [7]. Motivated by recent development in Bayesian brain theory that characterizes human or animal brains for adaptive perception and learning, we propose to use the free-energy principle as a unifying approach for adaptive waveform design and for sequential target parameter estimation. Given a probabilistic model of some data Y conditioned on parameters θ, we can show that by the Bayes' theory, the logarithm of the marginal likelihood (also called evidence) [8] can be written as

$$\log p(Y) = -F + KL[q(\theta)||p(\theta|Y)] \tag{1}$$

where the first term F in (1) is called the free energy

$$F = -\int q(\theta)\log\frac{p(Y,\theta)}{q(\theta)}d\theta \tag{2}$$

This work was supported in part by the U.S. Army Research Laboratory, the Office of Naval Research and the Army Research Office under grant no. W911NF-11-1-0160, the National Science Foundation under grant no. CMMI-1126008, and a subcontract from Carnegie Mellon University through the National Science Foundation under grant no. CNS-0930868.

and the second term $KL[q(\theta)||p(\theta|Y)]$ is the Kullback Liebler (KL) divergence from the density $q(\theta)$ to the posterior density $p(\theta|Y)$, which is defined as

$$KL[q(\theta)||p(\theta|Y)] = \int q(\theta)\log\frac{q(\theta)}{p(\theta|Y)}d\theta \tag{3}$$

It is readily to see from (1) and (3) the KL divergence is minimized when the density $q(\theta)$ is equal to the true posterior density $p(\theta|Y)$. Hence, minimizing the free energy F is equivalent to maximizing the model evidence $p(Y)$. This observation implies that, from the statistical inference standpoint, minimizing free energy is amount to choosing the approximate posterior $q(\theta)$ as close as possible to the true posterior. This becomes the underlying reason for the well known variational Bayesian (VB) approach in machine learning [9], signal processing [10, 11] and neuro-imaging data analysis [8, 12].

This paper builds upon our prior work in adaptive waveform design and sequential Bayesian estimation for cognitive radar [13, 3]. In [13] we developed a sequential Bayesian estimation algorithm that utilizes adaptive waveform methods to achieve reduced estimation error variance and faster convergence for the estimator. However, the design of the estimator and the design of adaptive waveform are based on two different metrics, i.e., the minimal mean squared error (MMSE) principle and the maximal mutual information (MI) principle, respectively. Hence, a unified framework for transmission waveform design and sequential target parameter estimation in the context of cognitive radar is lacking.

In this paper we consider the problem of sequential estimation of radar target using adaptive waveforms under the free energy principle. The cognitive radar iteratively transmits adaptive waveforms in response to the received radar measurement. The estimation problem and the waveform design problem are formulated by minimizing the free energy that is dependent of the radar target return, the transmission waveform, and the target parameter to be estimated. We will show that, using adaptive waveforms, the proposed method achieves faster convergence and lower estimation error variances compared with the non-adaptive waveform method. Furthermore, under the assumption of the Gaussian data model in the presence of a single unknown parameter, we show that the FE estimator is the same as the conventional Bayesian MMSE estimator and that the FE waveform design results in similar solution to the MI based waveform design.

2. SIGNAL MODEL

Here we consider a mono-static configuration where the radar antenna can switch between the transmit mode and the receive mode. The received backscattered signal can be written in the frequency domain as

$$Y_j(f_q) = H_j(f_q)S_i(f_q) + W_j(f_q) \tag{4}$$

where $S_i(f_q)$ is the frequency spectrum of the transmitted waveform with the energy constraint $\frac{1}{Q}\sum_{q=1}^{Q} S_i(f_q)S_i^*(f_q) \leq E_s$. $W_j(f_q)$ is the power spectrum of the additive Gaussian noise. $H_j(f_q)$ is the Fourier transform of the impulse response of the target. Here, i is the index for transmission cycle. Each cycle consists of $l = 1, \cdots L$ repeated transmission and reception. At i-th transmission cycle, a different waveform $S_i(f_q)$ is used while within each cycle, the same waveform $S_i(f_q)$ is used. The j-th (where $j = (i-1) \times l$) measurement vector is given by

$$\mathbf{y}_j = [Y_j(f_q), \cdots, Y_j(f_Q)], \ j = 1, \cdots, J \tag{5}$$

The probability models of the signals are given by

$$\begin{aligned} H_j(f_q) &\sim \mathcal{CN}\left(\mu_h(q), \sigma_h^2(q)\right) && (6)\\ Y_j(f_q) &\sim \mathcal{CN}\left(S(f_q)\mu_h(q), \sigma_y^2(q))\right) && (7)\\ W_j(f_q) &\sim \mathcal{CN}\left(0, \sigma_w^2(q)\right) && (8) \end{aligned}$$

For simplicity, here we assume that the mean $\mu_h(q)$ of the target response is unknown and needs to be estimated. The goal of the sequential estimation problem is two-fold: First, to obtain the estimate the mean of target response $\mu_h(q)$ in a Bayesian setting. Second, to design a waveform $S_i(f_q)$ such that the estimation process converges faster. Both design problems are solved using the minimal free energy principle.

3. SEQUENTIAL ESTIMATION BY FREE ENERGY PRINCIPLE

By the sequential Bayesian estimation theory [14, 15], we model the unknown $\mu_h(f_q)$ as a random process α_q with an initial prior Gaussian distribution $p_0(\alpha_q) = \mathcal{CN}\left(\alpha_0, \sigma_\alpha^2(q)\right)$. Let $r_J(\alpha_q)$ denote the recognition density of α_q at the J-th measurement. The recognition density is an approximation of the true density of the parameter of interest. When $J = 0$, we choose $r_J(\alpha_q) = p_0(\alpha_q)$. Ideally,

$$r_J(\alpha_q) \to \mu_h(q)\delta\left(\alpha_q - \mu_h(q)\right) \ \text{ as } J \to \infty. \tag{9}$$

Hence, the parameter estimation problem amounts to computing the recognition density $r_J(\alpha_q)$ so that the estimate at J-th measurement is

$$\widehat{\alpha}_q(J) = \int \alpha_q r_J(\alpha_q) d\alpha_q \tag{10}$$

Let $Y_{1:J}(f_q)$ denote the collection of measurements $\{Y_j(f_q)\}$, applying $r_J(\alpha_q)$ and $p(Y_{1:J}(f_q), \alpha_q)$ to (2), we obtain the expression for free energy at f_q as

$$\begin{aligned} F(q) &= -\int r_J(\alpha_q) \log \frac{p(Y_{1:J}(f_q), \alpha_q)}{r_J(\alpha_q)} d\alpha_q && (11)\\ &= KL[r_J(\alpha_q) \| p(\alpha_q|Y_{1:J}(f_q))] - \log p(Y_J(f_q)) \\ &\quad -\sum_{j=1}^{J-1} \log p(Y_j(f_q)) && (12) \end{aligned}$$

where the total free energy is $F = \sum_{q=1}^{Q} F(q)$. Note that the term $-\sum_{j=1}^{J-1} \log p(Y_j(f_q))$ in (12) is non-negative, we then re-define the free energy by discarding this term, which yields

$$F(q) = KL[r_J(\alpha_q) \| p(\alpha_q|Y_{1:J}(f_q))] - \log p(Y_J(f_q)) \tag{13}$$

Furthermore, noticing that $Y_J(f_q)$ is a random variable, we then define the expected free energy as

$$\begin{aligned} \widetilde{F}(q) &= \langle F(q) \rangle_{Y_J(f_q)} && (14)\\ &= H(Y_J(f_q)) + \langle KL[r_J(\alpha_q) \| p(\alpha_q|Y_{1:J}(f_q))] \rangle_{Y_J(f_q)} \end{aligned}$$

where the expectation is defined by $\langle x \rangle_Y = \int p(Y)xdY$ and the entropy is defined by

$$H(Y_J(f_q)) = -\int p(Y_J(f_q)) \log p(Y_J(f_q)) dY_J(f_q) \tag{15}$$

Thus, the expected free energy in (14) is expressed as the sum of the measurement entropy and the KL divergence between the recognition density and the posterior density. Furthermore, notice that the free energy is a function of recognition density $r_J(\alpha_q)$, the measurement $\mathbf{y}_J(f_q)$ and the energy spectrum density $\tau_i(f_q) = |S_i(f_q)|^2$ of the transmission waveform, the estimation problem is expressed as

$$r_J(\alpha_q) = \arg\min_{r_J(\alpha_q)} \sum_{q=1}^{Q} \widetilde{F}\left(r_J(\alpha_q), \mathbf{y}_J(f_q), \tau_i(f_q)\right) \tag{16}$$

Note that the entropy term in (14) is independent of the recognition density. Hence, (16) is simplified to

$$r_J(\alpha_q) = \arg\min_{r_J(\alpha_q)} \sum_{q=1}^{Q} KL\left[r_J\left(\alpha_q\right) || p\left(\alpha_q|Y_{1:J}(f_q)\right)\right] \tag{17}$$

Next, the non-negative divergence term depends on recognition density $r_J(\alpha_q)$ and equals zero if $r_J(\alpha_q)$ equals the posterior density. Hence, the total free energy in the system is minimized if the recognition density is equal to the posterior conditional density, which is given by

$$r_J\left(\alpha_q\right) = p(\alpha_q|Y_{1:J}(f_q)) = \frac{p(\alpha_q, Y_{1:J}(f_q))}{p\left(Y_{1:J}(f_q)\right)} \tag{18}$$

Since the relation between the parameter α_q and $Y_J(f_q)$ is linear, the prior distribution of α_q is Gaussian, and both the conditional densities of $Y_J(f_q)$ and $H_J(f_q)$ are Gaussian, we can easily prove that the posterior density of α_q is also Gaussian, which takes the form

$$\alpha_q|Y_{1:J}(f_q) \sim \mathcal{CN}\left(\mu_{\alpha_q}(J), \sigma^2_{\alpha_q|Y_{1:J}(f_q)}\right) \tag{19}$$

Next, we define $\sigma^2_{\alpha_q,J} = \int(\alpha_q - \widehat{\alpha}_q(J))^2 r_J(\alpha_q)d\alpha_q$, where $\widehat{\alpha}_q(J)$ is defined in (10). By (18), we obtain the estimate and the error variance $\sigma^2_{\alpha_q|Y_{1:J}}(f_q)$ (denoted by $\nu_J(q)$) as follows:

$$\begin{aligned} \widehat{\alpha}_q(J) &= \mu_{\alpha_q}(J) && (20)\\ &= \nu_J(q)\left(\frac{\mu_\alpha(J-1)}{\nu_{J-1}(q)} + \frac{S_i^*(f_q)Y_J(f_q)}{\sigma_w^2 + \tau_i(q)\sigma_h^2(q)}\right) \\ \frac{1}{\nu_J(q)} &= \frac{1}{\nu_{J-1}(q)} + \frac{\tau_i(q)}{\sigma_w^2 + \tau_i(q)\sigma_h^2(q)} && (21)\end{aligned}$$

Eqn. (18) shows that, for a single parameter, the minimal FE principle is equivalent to the MMSE method for sequential Bayesian estimation.

4. WAVEFORM DESIGN BY FREE-ENERGY PRINCIPLE

We define the $(J-1)$-th prior distribution for the random process α_q

$$P_{J-1}(\alpha_q) = p_0(\alpha_q)\prod_{j=1}^{J-1} p(Y_j(f_q)|\alpha_q)/\prod_{j=1}^{J-1} p(Y_j(f_q)) \tag{22}$$

The expected free energy defined in (14) can be re-written as

$$\begin{aligned}\widetilde{F}(r_J(\alpha_q), Y_J(f_q), \tau_i(q)) &= KL\left[r_J(\alpha_q)\,||p_{J-1}(\alpha_q)\right] \\ &\quad - \langle \log p\left(Y_J(f_q)|\alpha_q\right)\rangle_{r_J(\alpha_q),p(Y_J(f_q))} \qquad (23)\end{aligned}$$

By the Bayesian brain theory, we define *complexity* as the divergence between recognition density and prior density and *accuracy* as the expected likelihood over the parameter and measurement space. Hence the free energy in (23) can be interpreted as difference between complexity (first term) and accuracy (second term). The waveform design problem is formulated as the free energy minimization problem given below:

$$\begin{aligned}\tau_i(q) = \arg\min_{\tau_i(q)} \quad & \textstyle\sum_{q=1}^{Q} \widetilde{F}(r_J(\alpha_q), Y_J(f_q), \tau_i(q)) \\ \text{subject to} \quad & \tfrac{1}{Q}\textstyle\sum_{q=1}^{Q}\tau_i(q) \le E_s \qquad (24)\end{aligned}$$

Notice from (23) that only the second term (accuracy) depends on the waveform energy spectrum $\tau_i(q)$, the minimization problem in (24) can be re-formulated as

$$\begin{aligned}\tau_i(q) &= \arg\max_{\tau_i(q)} \sum_{q=1}^{Q} \langle \log p\left(Y_J(f_q)|\alpha_q\right)\rangle_{r_J(\alpha_q),p(Y_J(f_q))} \\ &\qquad \text{subject to } \frac{1}{Q}\sum_{q=1}^{Q}\tau_i(q) \le E_s \qquad (25)\end{aligned}$$

Note that $Y_J(f_q)|\alpha_q \sim \mathcal{CN}\left(S_i(f_q)\alpha_q, \sigma_w^2(q)\right)$, we obtain

$$\begin{aligned} acc_q(J) &\triangleq \langle \log p\left(Y_J(f_q)|\alpha_q\right)\rangle_{r_J(\alpha_q),p(Y_J(f_q))} && (26)\\ &= -\log\pi - \log\left(\sigma_w^2(f_q) + \tau_i(q)\sigma_h^2(f_q)\right) \\ &\quad - \frac{\left\langle |Y_J(f_q) - S_i(f_q)\alpha_q|^2\right\rangle_{r_J(\alpha_q),p(Y_J(f_q))}}{\sigma_w^2(f_q) + \tau_i(q)\sigma_h^2(f_q)} && (27)\end{aligned}$$

Furthermore, by an approximation of $\langle\alpha_q\rangle_{r_J(\alpha_q)} = \mu_{\alpha_q}(J)$ and $\langle|\alpha_q|^2\rangle_{r_J(\alpha_q)} = \nu_J(q) + |\mu_{\alpha_q}(J)|^2$, we obtain

$$\begin{aligned}\langle |Y_J(f_q) - S_i(f_q)\alpha_q|^2\rangle_{r_J(\alpha_q),p(Y_J(f_q))} &= \sigma_w^2(f_q) \\ &\quad +\tau_i(q)\left(\sigma_h^2(f_q) + \nu_{J-1}(q)\right) \qquad (28)\end{aligned}$$

Therefore, (27) becomes

$$\begin{aligned} acc_q(J) &= -\log\pi - \log\left(\sigma_w^2(f_q) + \tau_i(q)\sigma_h^2(f_q)\right) \\ &\quad - \frac{\sigma_w^2(f_q) + \tau_i(q)\left(\sigma_h^2(f_q) + \nu_{J-1}(q)\right)}{\sigma_w^2(f_q) + \tau_i(q)\sigma_h^2(f_q)} \qquad (29)\end{aligned}$$

Using (29), the optimization problem in (25) for designing $\tau_i(q)$ can be solved by the Lagrange form (ignoring terms that are independent of $\tau_i(q)$) as

$$L(\tau_i,\lambda) = \sum_{q=1}^{Q}\left[\log\left(\sigma_w^2(f_q) + \tau_i(q)\sigma_h^2(f_q)\right) + \frac{\sigma_w^2(f_q) + \tau_i(q)\left(\sigma_h^2(f_q) + \nu_{J-1}(q)\right)}{\sigma_w^2(f_q) + \tau_i(q)\sigma_h^2(f_q)}\right] + \lambda\left(\sum_{q=1}^{Q}\tau_i(q) - E_s\right)$$

By differentiating the above Lagrange function with respect to $\tau_i(q)$, we obtain

$$\begin{aligned}&\sigma_h^2(f_q)\left(\sigma_w^2(f_q) + \tau_i(q)\sigma_h^2(f_q)\right) + \nu_{J-1}(q) \\ &\sigma_w^2(f_q) + \lambda\left(\sigma_w^2(f_q) + \tau_i(q)\sigma_h^2(f_q)\right)^2 = 0 \qquad (30)\end{aligned}$$

Noticing that (30) is a quadratic equation for $\tau_i(q)$, we solve for the roots of the quadratic equation and select only the non-negative root, which yields

$$\tau_i^{opt}(q) = \left[\frac{\eta}{2}\left(1 + \sqrt{1 + \frac{4\sigma_w^2(f_q)\nu_{J-1}(q)}{\eta\sigma_h^4(f_q)}}\right) - \frac{\sigma_w^2(f_q)}{\sigma_h^2(f_q)}\right]^+ \tag{31}$$

It can be observed that this is a water-filling type solution where $\eta = -\frac{1}{\lambda}$ is the constant water-level. The value of η can be evaluated by inserting the solution (31) by the energy constraint.

To benchmark the waveform design performance, we consider the well known maximal mutual information (MI) method by computing the MI between α_q and $Y_J(f_q)$ as

$$I(\alpha_q, Y_j(f_q)) = H(\alpha_q) - H(\alpha_q|Y_j(f_q)) \tag{32}$$

Since the pdfs of α_q and $\alpha_q|Y_J(f_q)$ in our problem are Gaussian, (32) can be re-written as

$$I(\alpha_q, Y_J(f_q)) = \log\frac{\nu_{J-1}(q)}{\nu_J(q)} \tag{33}$$

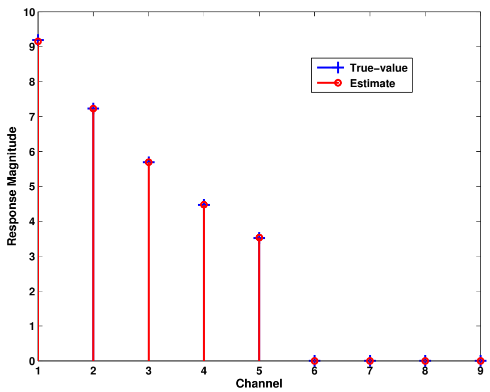

Fig. 1. Estimation of mean target response

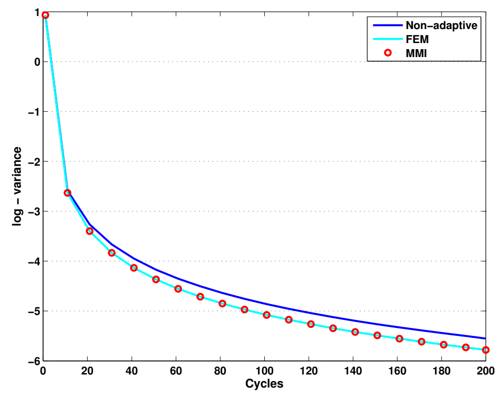

Fig. 2. Comparison of variance performance for adaptive vs. non-adaptive methods

The waveform that minimizes the inverse of this variance-ratio maximizes $I(\alpha_q, Y_J(f_q))$. Thus, the maximum mutual information waveform design is expressed as follows

$$\begin{aligned} \tau_i(q) = \arg\min_{\tau_i(q)} \quad & \sum_{q=1}^{Q} \frac{\nu_J(q)}{\nu_{J-1}(q)} \\ \text{subject to} \quad & \frac{1}{Q}\sum_{q=1}^{Q} \tau_i(q) \leq E_s \end{aligned} \tag{34}$$

This optimization problem is reformulated as a Lagrange function and a partial derivative with respect to $\tau_i(q)$ is evaluated, which results in the following quadratic equation:

$$\lambda(\sigma_w^2(f_q) + \tau_i(q)(\sigma_h^2(f_q) + \nu_{J-1}(q)))^2 = \sigma_w^2(f_q)\nu_{J-1}(q)$$

Since the waveform energy should be non-negative, only one solution exists, which can be re-written as a water-filling solution by using $\eta = \frac{1}{\sqrt{\lambda}}$, i.e.,

$$\tau_i^{opt}(q) = \left[\frac{\eta\sqrt{\sigma_w^2(f_q)\nu_{J-1}(q)} - \sigma_w^2(f_q)}{\sigma_h^2(f_q) + \nu_{J-1}(q)}\right]^+ \tag{35}$$

Eqns. (31) and (35) show that the FE method and the MI method lead to similar water-filling solutions.

5. NUMERICAL RESULTS

In this section, we present the numerical simulations to show that the adaptive waveforms in equations (31) and (35) result in same performance in terms of convergence of error variance $\nu_J(q)$. The simulation is setup as follows. The spectral variance of the target response with nine channels ($Q = 9$) is modeled by $\sigma_h^2(f_q) = \beta_1 + \beta_2 \exp -((f_q - f_c)/f_c)^2)$ with $f_c = f_{\lfloor \frac{Q}{2} \rfloor}$, where the values of β_1 and β_2 are assumed to be 0.4 and 0.6, respectively. The true value $\mu_h(q)$ is assumed to be an exponential function of f_q as $\mu_h(q) = a_0 \exp(-((f_q - f_c)/f_c)^2)$ where the peak value is $a_0 = 2 + 4i$. The total number of cycles is 200, the number of snapshots taken in each cycle is $L = 10$. Hence, the total number of iterations is $J = 2000$. At each q the target response is generated using Gaussian statistics. The radar measurements are generated (simulated) using the model in (4). In the non-adaptive case, the waveform is fixed as $S(f_q) = \sqrt{E_s}$. In the adaptive case, the waveforms are evaluated as $S_i(f_q) = \sqrt{\tau_i(q)}$ from the FE and MI waveforms, respectively. From the measurements, the posterior pdf of α_q is sequentially evaluated by (19) and (20). Fig. 1 shows the estimate of the magnitude of the target response overlayed on the true values. It can be seen the estimation process is successful. Also note that the channels $6 \sim 9$ do not have a significant target response in terms of SNR gain. The SNR gain is given by $SNRG_J(q) = \frac{|Y_j(f_q)|^2 - \sigma_w^2(f_q)}{\tau_i(q)}$. As illustrated in Table 1 these channels are characterized based on their SNR gain which is less than 5% of the maximum SNR gain (in terms of dB, the cutoff is $19 - 13 = 6\ dB$). Fig. 2 illustrates the behaviors of error variance $\nu_J(q)$ using fixed waveforms and adaptive waveforms designed by the FE and MI methods. The plots show that the adaptive method converges faster than the non-adaptive method as the number of iterations J increases. The performance for the FE and MI methods are the same.

Table 1. SNR gain at each channel

Channel (q)	1	2	3	4	5	$6 \sim 9$
SNR Gain (dB)	19	17	15	13	11	≤ 2

6. CONCLUSION

In this paper, we demonstrate the use of free energy as an unifying framework for both parameter density estimation of a single parameter and adaptive waveform design for faster convergence. For the problem of single parameter estimation under the Gaussian signal model, the FE approach is same as the Bayesian estimation. For waveform design, minimization of free energy gives rise to a water filling solution that can be iteratively updated. Hence, the free energy principle enables efficient and tractable computational schemes for parameter estimation applicable to cognitive radar system design.

7. REFERENCES

[1] S. Haykin, "Cognitive radar networks," in *IEEE International Workshop on omputational Advances in Multi-Sensor Adaptive Processing, 2005*, December 2005, pp. 1–3.

[2] M. R. Bell, "Information theory and radar waveform design," *IEEE Transactions on Information Theory*, vol. 39, no. 5, pp. 1578–1597, September 1993.

[3] Y. Jin, J. M. F. Moura, and N. O' Donoughue, "Time reversal in multiple-input multiple-output radar," *IEEE Journal of Selected Topics on Signal Processing*, vol. 4, no. 1, pp. 210–225, Feburary 2010.

[4] N. A. Goodman, P. R. Venkata, and M. A. Neifeld, "Adaptive waveform design and sequential hypothesis testing for target recognition with active sensors," *IEEE Journal of Selected Topics in Signal Processing*, vol. 1, no. 1, pp. 105–113, June 2007.

[5] J. Bae and N. A. Goodman, "Automatic target recognition with unknown orientation and adaptive waveforms," in *IEEE Radar Conference (RADAR)*. May 2011, pp. 1000–1005, IEEE.

[6] P. Chavali and A. Nehorai, "Scheduling and Power Allocation in a Cognitive Radar Network for Multiple-Target Tracking," , *IEEE Transactions on Signal Processing*, vol. 60, no. 2, pp. 715–729, 2012.

[7] U. Gunturkun, "Toward the development of radar scene analyzer for cognitive radar," *IEEE Journal of Oceanic Engineering*, vol. 35, no. 2, pp. 303–313, April 2010.

[8] W. D. Penny, S. J. Kiebel, and K. J. Friston, "Variational Bayesian inference for fMRI time series," *NeuroImage*, vol. 19, no. 3, pp. 727–741, 2003.

[9] C. Peterson and J. Anderson, "A mean field theory learning algorithm for neural networks," *Complex Systems*, vol. 1, pp. 995–1019, 1987.

[10] T. S. Jaakkola and M. I. Jordan, "A variational approach to Bayesian logistic regression models and their extensions," Tech. Rep. 9702, MIT Computational Cognitive Science, January 1997.

[11] J. Winn and C. Bishop, "Variational message passing," *Journal of Machine Learning Research*, vol. 6, pp. 661–694, 2005.

[12] K. Friston, "The free-energy principle: a rough guide to the brain," *Trends in Cognitive Science*, vol. 13, no. 7, pp. 293–301, June 2009.

[13] A. Turlapaty and Y. Jin, "Sequential Bayesian parameter estimation by cognitive multi-antenna arrays," (submitted for publication).

[14] M. Alvo, "Bayesian sequential estimation," *The Annals of Statistics*, vol. 5, no. 5, pp. 955–968, September 1977.

[15] J. V. Candy, *Bayesian Signal Processing*, Wiley-IEEE Press, Hoboken, NJ, 2009.