A Stochastic Newton-Raphson Method with Noisy Function Measurements

Khaled Kamal Saab, *Member, IEEE*, and Samer Said Saab, Jr.

***Abstract*—This letter shows that traditional Newton-Raphson (NR) method cannot achieve zero-convergence in presence of additive noise without adding a multiplicative gain. Furthermore, this gain needs to converge to zero. This article proposes a novel recursive algorithm providing optimal iterative-varying gains associated with the NR method. The development of the proposed optimal algorithm is based on minimizing a stochastic performance index. The estimation error covariance matrix is shown to converge to zero for linearized functions while considering additive zero-mean white noise. In addition, the proposed approach is capable of overcoming common drawbacks associated with the traditional NR method. Simulation results are included to illustrate the performance capabilities of the proposed algorithm. We show that the proposed recursive algorithm provides significant improvement over the traditional NR method.**

***Index Terms*—Newton-Raphson method, noisy function measurements, stochastic optimization.**

I. Introduction

MANY engineering problems make use of an optimization setting where only noisy estimates of the objective function are available. Localization problems based on triangulation consist of two methods: lateration and angulation. Lateration methods estimate the target location by measuring its distances from multiple reference points whereas angulation, measure angles relative to several reference points. The needed measurements are normally noisy where the noise depends on the technology under consideration. For example, lateration methods can be based on noisy received signal strength measurements, see, e.g., [1]–[3]. Other classes of applications include design of composite materials [4], target tracking [5], recovery of sparse and compressible signals [6], image sampling and color analysis [7], and acoustic wave propagation in turbulent fluids [8]. Deterministic optimization methods are frequently implemented in such noisy-based applications. However, stochastic optimization techniques provide an effective approach in the presence of noisy measurements.

While considering gradient-based descent approach, many methods have been proposed to improve the choices of the step size such as the optimal version of Robbins-Monro (RM) algorithm [9], an accelerated RM algorithm [10], [11], and an accelerated version of Kesten algorithm [12]. Stable constants are introduced to the step size in order to improve algorithm stability [13]. An efficient approach for achieving a second order adaptive algorithm is presented in [14], where two parallel recursions are implemented for estimating the solution using NR algorithm, and the other for estimating the Hessian matrix. Stochastic gradient algorithm over a subset of the parameters while the rest are held fixed is shown to be an effective approach [15]. The above review is by no means exhaustive of the type of problems for which noisy measurements are considered.

Although gradient-based descent methods are shown to be rather effective, they generally suffer from slow convergence when contrasted with the NR method. In this letter we investigate the implementation of the NR method that attempts to find zeroes of functions which cannot be computed directly, but only estimated from noisy measurements of the functions. This letter shows that the traditional NR method cannot guarantee zero convergence in presence of additive measurement noise. Consequently, we consider adding a multiplicative gain matrix to the NR correction factor. In addition, this letter proposes a novel recursive algorithm providing optimal iterative-varying gain associated with the NR method. The proposed recursive algorithm utilizes the statistical measurement error model in order to construct the iterative-varying gain matrix while minimizing the variance of errors. Analytical convergence results are also provided. The analytical results show that the error covariance matrix converges to zero, which show the capability of the proposed algorithm rejecting measurement noise. Furthermore, two simple examples are provided illustrating how the proposed method is capable of overcoming some of the NR common drawbacks. Simulation results are also included to illustrate the performance abilities of the proposed algorithm and its advantages over the traditional NR method.

II. Problem Formulation

A zero-finding problem gives M functional relations to be zeroed, that is,

$$f(z) = 0 \tag{1}$$

where $z \in \mathbb{R}^N$, and $f(.) \in \mathbb{R}^M$. The assumption in the zero-finding setting is that $f(.)$ is not available directly, but must be estimated through a noisy estimate of $f(.)$, $\hat{f}$. In this letter, we consider noisy observations of $f(.)$ consisting of additive noise, that is, $\hat{f} = f(.) + \epsilon$, and consider unconstrained optimization

Manuscript received November 20, 2015; accepted December 13, 2015. Date of publication December 23, 2015; date of current version February 05, 2016. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Martin Ulmke.

K. K. Saab is with the Department of Electrical and Computer Engineering at Georgia Institute of Technology, Atlanta, GA 30332 USA (e-mail: ksaab3@gatech.edu).

S. S. Saab, Jr. is with the Department of Electrical and Computer Engineering at Pennsylvania State University, State College, PA 16801 USA (e-mail: sys5880@psu.edu).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/LSP.2015.2511456

and the case where during each iteration, one measurement of $f(.)$ is available. We consider the following setting at iterative instant, k:

$$\hat{f}(.) = f(.) + g(k)v(k) \tag{2}$$

where $v(k) \in \mathbb{R}^M$ a zero-mean white random process, $g(k) \in \mathbb{R}^{M\times M}$ is a deterministic function. In the neighborhood of $\hat{z} \in \mathbb{R}^N$, while assuming that the elements of f are continuously differentiable, f can be expanded in Taylor series as follows:

$$f(\hat{z} + \Delta z) = f(\hat{z}) + J(k)\Delta z + O\left(\Delta z^2\right)$$

where the elements of the Jacobian matrix, $J(.) \in \mathbb{R}^{M\times N}$, are defined as $J_{ji}(k) \triangleq \frac{\partial f_j}{\partial z_i}|_{\hat{z}=\hat{z}(k)}$. In what follows, we neglect terms of order Δz^2 and higher leading to

$$f(\hat{z} + \Delta z) = f(\hat{z}) + J(k)\Delta z \tag{3}$$

Problem Statement: Assume there exists a z such that $f(z) = 0$. Given erroneous values of $f(z)$, $\hat{f}(\hat{z})$, at each time instant, develop a recursive algorithm such that the variance of $(z - \hat{z})$ is minimized at each time instant.

III. Proposed Optimal Recursive Algorithm, Convergence and Characteristics

This section addresses the problem under consideration. In particular, the proposed stochastic algorithm leading to optimal gains is developed based on a linearized set of functions (3), and convergence results are included. Two examples are also included in order to show how the proposed approach can be used to overcome the NR common drawbacks. Similar to Newton-Raphson method, we set $f(\hat{z}(k) + \Delta z) \equiv 0$, and we also set $g(k)v(k) \equiv 0$. Thus, it holds

$$J(k)\Delta z = -f(\hat{z}(k)) \tag{4}$$

We assume that $J(k)$ is full-column rank. If a solution to the problem $f(z) = 0$ exists, then it is adequate to multiply both sides of (4) by Moore–Penrose pseudoinverse of $J(k)$, $J^\dagger(k) = [J^T(k)J(k)]^{-1}J^T(k)$. Consequently, (4) yields

$$\Delta z = -J^\dagger(k)f(z) \tag{5}$$

In this work, we refer to the following iterative method as NR:

$$\hat{z}(k+1) = \hat{z}(k) - J^\dagger(k)\hat{f}(\hat{z}(k)) \tag{6}$$

where we set $\Delta z \equiv \hat{z}(k+1) - \hat{z}(k)$. However, this solution may not be optimal in presence of errors or measurement noise, $v(k)$. In order to possibly find a more suitable approach, we add a multiplicative gain $K(k) \in \mathbb{R}^{N\times N}$, in this fashion:

$$\Delta z = -K(k)J^\dagger(k)\hat{f}(\hat{z}) \tag{7}$$

Again set $\Delta z \equiv \hat{z}(k+1) - \hat{z}(k)$ to obtain

$$\hat{z}(k+1) = \hat{z}(k) - K(k)J^\dagger(k)\hat{f}(\hat{z}(k)) \tag{8}$$

Remark 1: Equation (8) resembles the corrector step of a Kalman filter. However, it is basically quite different. In particular, a Kalman filter is based on state space of dynamical systems. For a linear discrete-time invariant stochastic system and in absence of the input signal, the state-space equation becomes $z(k+1) = Az(k) + Gw(k)$, where $z(.)$ represents the state vector and $w(.)$ represents noise. This article is concerned with finding the zero of $f(z)$, that is, finding z, which is most likely different than zero, such that $f(z) = 0$, that is, in general $z \neq 0$. Assume $f(z)$ is a linear function of z. If $f(z)$ were to be considered as $f(z) = Az(k)$, then in absence of noise, the state space equation would become $z(k+1) = Az(k) = f(z) = 0$. The latter leads to $z(k+1) = 0, \forall k$, which cannot be applied to the problem under study as it stands. On the other hand, the work in [16] presents an iterated Kalman filter, which adopts the NR iterative optimization steps, however, the problem addresses state estimation of a nonlinear stochastic discrete-time system, which is different than the problem addressed in this letter.

Theorem 1: Consider the linear vector function given in (3) and the method proposed in (8). Assume that there exists a z such that $f(z) = 0$, and the Jacobian matrix is full-column rank, $k \geq 0$. The gain $K(k)$ that minimizes the mean-square of $\delta z(k) \triangleq z - \hat{z}(k)$ at each kth instant is given in the following recursive formulas for all $k > 0$,

$$K(k) = P(k)(P(k) + F(k)R(k)F^T(k))^{-1} \tag{9}$$

$$P(k+1) = (I - K(k))P(k) \tag{10}$$

where $P(k) \triangleq E[\delta z(k)\delta z^T(k)]$, $F(k) \triangleq J^\dagger(k)g(k)$ and $R(k) \triangleq E[v(k)v^T(k)]$.

Proof of Theorem 1: Inserting (2) in (8), we obtain $\hat{z}(k+1) = \hat{z}(k) - K(k)J^\dagger(k)f(\hat{z}(k)) - K(k)J^\dagger(k)g(k)v(k)$ From (3), we have $f(\hat{z} + \delta z) = f(\hat{z}) + J(k)\delta z$. In what follows, we use $\delta z(k) = z - \hat{z}(k)$ to obtain $f(z) = f(\hat{z}(k)) + J(k)(z - \hat{z}(k))$. But $f(z) = 0$, thus $f(\hat{z}(k)) = -J(k)(z - \hat{z}(k))$, or $\hat{z}(k+1) = \hat{z}(k) + K(k)(z - \hat{z}(k)) - K(k)J^\dagger(k)g(k)v(k)$.

Subtracting both sides of this equation from z and rearranging terms yield

$$\delta z(k+1) = (I - K(k))\delta z(k) + KJ^\dagger(k)g(k)v(k) \tag{11}$$

For compactness, we denote by $P_{k+1} \triangleq P(k+1)$, $P_k \triangleq P(k)$, $F \triangleq J^\dagger(k)g(k)$, $K \triangleq K(k)$ and $R \triangleq E[v(k)v^T(k)]$. Taking the covariance on both sides of (11) and observing that $E[\delta z(k)v^T(k)] = 0$, we have

$$P_{k+1} = (I - K)P_k(I - K)^T + KFRF^TK^T \tag{12}$$

Expending terms of (12) yields

$$P_{k+1} = P_k + KP_kK^T - KP_k - P_kK^T + KFRF^TK^T \tag{13}$$

To minimize $\mathrm{tr}(P_{k+1})$, where tr is the trace operator, with respect to K, we set $\frac{\partial \mathrm{tr}(P_{k+1})}{\partial K} \equiv 0$ at each instant, $\frac{\partial \mathrm{tr}(P_{k+1})}{\partial K} = +2KP_k - 2P_k + 2KFRF^T \equiv 0$. Therefore, $K = P_k(P_k + FRF^T)^{-1}$. Inserting this optimal value of K

in (13) and collecting terms lead to $P_{k+1} = P_k - KP_k + P_kK^T + P_k(P_k + FRF^T)^{-1}(P_k + FRF^T)K^T$ Cancelling then collecting terms yield (10). ■

Proposition 1: Assume that $R > 0$ and F is full rank. If $\lim_{k\to\infty} P(k) = 0$, then it is necessary to have $\lim_{k\to\infty} K(k) = 0$.

Proof of Proposition 1: We first note that $FRF^T > 0$. Since $(I - K(k))P(k)(I - K(k))^T \geq 0$ for $K(k) \neq 0$ (e.g., in NR, $K = I$), then (12) implies $\lim_{k\to\infty} P(k) \neq 0$, $\lim_{k\to\infty} P(k) = 0$ only if $\lim_{k\to\infty} K(k) = 0$. ■

Remark 2: In presence of measurement errors, the traditional NR method cannot lead to zero-error convergence. $\lim_{k\to\infty} K(k) = 0$ is a necessary condition. The latter requires an iterative-varying gain.

Theorem 2: Consider the recursive algorithm presented in (9) and (10). If $P(0) > 0$ and $F(k)R(k)F^T(k) > 0$, $\forall k \geq 0$, then $0 < \lambda(I - K(k)) < 1$, $\lim_{k\to\infty} P(k) = 0$, and $\lim_{k\to\infty} K(k) = 0$, where $\lambda(M)$ denotes the eigenvalues of M.

Proof of Theorem 2: Consider $I - K = I - P_k(P_k + FRF^T)^{-1}$. We first use an induction argument to show that $0 < \lambda(I - K) < 1$. For $k = 0$, we have both P_k and FRF^T positive definite, and $P_k(P_k + FRF^T)^{-1} = (I + FRF^TP_k^{-1})^{-1}$. Since $FRF^T > 0$ and $P_k^{-1} > 0$, then $\lambda(FRF^TP_k^{-1}) > 0$. Thus, $\lambda(I + FRF^TP_k^{-1}) > 1$. Therefore, $0 < \lambda(P_k(P_k + FRF^T)^{-1}) < 1$ and $0 < \lambda(I - P_k(P_k + FRF^T)^{-1}) < 1$. Consequently, $I - K > 0$ and $\lambda(P_1) > 0$. Since the covariance matrix P_1 is symmetric then $P_1 > 0$. There existsa multiplicative norm such that $\|I - K\| < 1$, thus $\|P_1\| < \|P_0\|$. We use similar arguments for any $k > 0$ to show that $0 < \lambda(I - P_k(P_k + FRF^T)^{-1}) < 1$ and $\|P_{k+1}\| < \|P_k\|$. Therefore, $\|P_k\|$ is bounded $\forall k \geq 0$ and for $k \to \infty$. If $\lim_{k\to\infty} P_k = 0$, this ends the proof, else $\lim_{k\to\infty} \|I - K\| < 1$, hence $\lim_{k\to\infty} P_k = 0$ and from (9) $\lim_{k\to\infty} K(k) = 0$. ■

Remark 3: In absence of measurement errors, $R = 0$, thus, (9) implies $K = I$, (6) and (7) become identical.

Remark 4: The results of Theorems 1 and 2 assume a linear system of functions without considering modeling errors due to linearization. Consequently, $\lim_{k\to\infty} P_k$ may not be zero. For example, if errors due to linearization are added to (3) and modelled as additive zero-mean white noise with covariance $Q_k \geq 0$, then it can be shown that the optimal gain remains $K_k = P_k(P_k + FRF^T)^{-1}$ and the associated covariance becomes $P_{k+1} = (I - K)P_k + Q_k$, which is always bounded; however, $\lim_{k\to\infty} P_k \neq 0$. Such a scenario is not considered in the proposed algorithm since the modelling part of such errors is function specific. In some cases of nonlinearities, P_k and K_k may converge to zero too early. In order to remedy this problem, we reset $P_{k+1} \equiv (I - K)P_k + Q_k$ for an arbitrary Q_k after a couple of iterations.

Remark 5: Although the implementation of the proposed method is more involved than NR, it is capable of overcoming some of the NR drawbacks. For example, at a stationary point or whenever the Jacobian matrix is not full-column rank, NR fails. The proposed method comes with some tuning flexibility, in particular, specific selection of P_0 and R, may overcome such problems. We present the following two examples, while disregarding measurement errors.

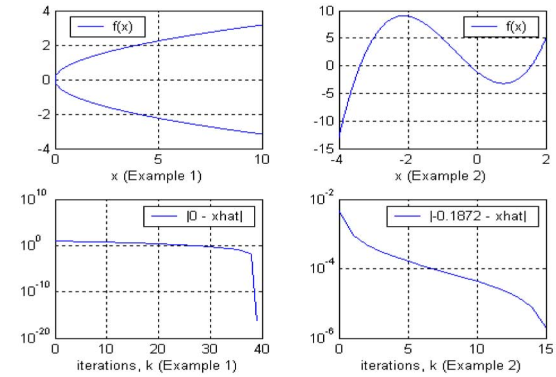

Fig. 1. Examples 1 and 2: $f(x)$ and $|x - \hat{x}(k)|$.

Example 1: Consider $\sqrt{x} = 0$ with initial value $x_0 = 10$. The NR roots oscillate between $x = 10$ and $x = -10$. On the other hand, when we set $P(0) = R = 100$, such large values indicate that initial guess and measurements are unreliable. The latter forces the proposed algorithm in taking smaller and smaller steps while searching for a solution. At the 40th iteration, the solution reaches $x = -5 \times 10^{-17} \cong 0$, see Fig. 1.

Example 2: Consider $x^3 + 2x^2 - 5x - 1 = 0$ with initial value $x_0 = -1.9$, which is considered close to the stationary point at $x = -2.12$. The iteration of NR method gives the root $x = 1.576$ instead of the closer root at $x = -0.1872$. The latter shows that NR can fail when initial value is close to a stationary point. Whereas by employing the proposed method, with a small value of $P(0)$, e.g., $P(0) = 0.5$ and $R = 3$, forces the algorithm to look for a solution around the initial guess. In fact, after one iteration we obtain $x = -0.1827$ and after four iterations the solution converges to $x = -0.187$, see Fig. 1.

IV. Example 3

In this example, we compare the performance of proposed scheme with the traditional NR method. In order to illustrate the performance capabilities of the proposed approach, we consider a set of nonlinear functions corrupted with measurement noise. The functions, without noise (1), are given by

$$f(x) = \begin{cases} x_1^3 + 2x_1^2 + x_2^2 + x_2 - 5 = 0 \\ \frac{1}{2}x_1^2 + 2x_1 - \frac{1}{2}x_2^2 - 5x_2 + x_1x_2 + 2 = 0 \\ -x_1^3 + \frac{1}{2}x_2^2 - (x_1x_2)^2 + 1.5 = 0 \end{cases} \quad (14)$$

The exact solution of (14) is simply $x_1 = x_2 = 1$. According to the model in (3), the additive measurement noise is given by

$$g(k)v(k) = \begin{bmatrix} -1 - \hat{x}_2(k) & 0 & 0 \\ 0 & 1 + \hat{x}_1(k) & 0 \\ 0 & 0 & -\hat{x}_1(k) - \hat{x}_2(k) \end{bmatrix} \begin{bmatrix} v_1(k) \\ v_2(k) \\ v_3(k) \end{bmatrix}$$

where $\hat{x}_i(k), i \in \{1, 2\}$ represents the estimate and $v_j(k) \in \aleph(0, \sigma), j \in \{1, 2, 3\}$ is zero-mean white Gaussian noise with standard deviation equals to σ. In what follows, we compare the performance of the NR method (6) with the proposed stochastic NR (StNR) presented in (8) while using the proposed recursive algorithm (9) and (10) in order to update the gain $K(k)$ and the error covariance matrix $P(k)$.

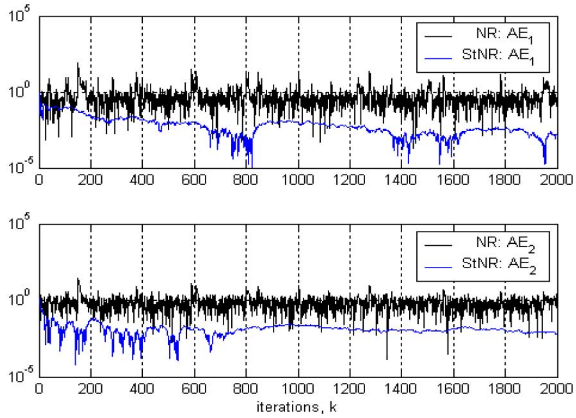

Fig. 2. Performance of NR versus StNR with $\sigma = 2$.

TABLE I
NR VS. STNR: ABSOLUTE ERRORS AT $k = 2000$

	$AE_1^{k=2000}$		$AE_2^{k=2000}$	
σ	NR	StNR	NR	StNR
0.1	0.0162	0.0018	0.0259	0.0013
1	0.1639	0.0038	0.803	0.0066
2	0.9199	0.0076	1.0422	0.0118
5	2.5138	0.0203	2.0609	0.0344
10	4.0632	0.0440	3.6393	0.0643

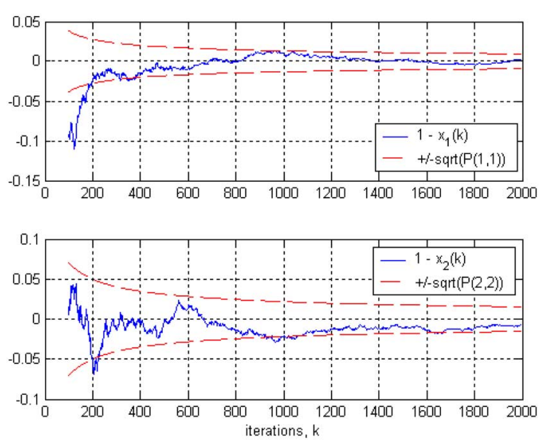

Fig. 3. Errors (in Blue) associated with StNR with $\sigma = 2, 100 \leq k \leq 2000$. The dashed plots (in Red) are the square roots of the diagonal elements of $P(k)$.

We set $\hat{x}_i(0) \equiv 2, i \in \{1, 2\}$ for both methods and $P(0) = 4I$ and $R(k) = \sigma^2 I, \forall k$. Fig. 2 shows the absolute errors, $\text{AE}_i \triangleq |1 - \hat{x}_i(k)|, i \in \{1, 2\}$, for $\sigma = 2$. One hundred independent runs, with 2,000 iterations per run, are also conducted for different values of σ. We extract the final value for each run and then take the average of absolute errors over the 100 runs, $\text{AE}_i^{k=2000} = \text{AVG}_{100\text{runs}}|1 - \hat{x}_i(2,000)|, i \in \{1, 2\}$. The values of $\text{AE}_i^{k=2000}$ for different values of σ are listed in Table I. Fig. 3 shows the errors, corresponding to StNR, superimposed with the relevant estimates of error standard deviations extracted from the diagonal elements of $P(k)$. Whereas Fig. 4 shows the elements of gain $K(k)$. While examining Table I and Fig. 2 to Fig. 4, the following can be concluded:

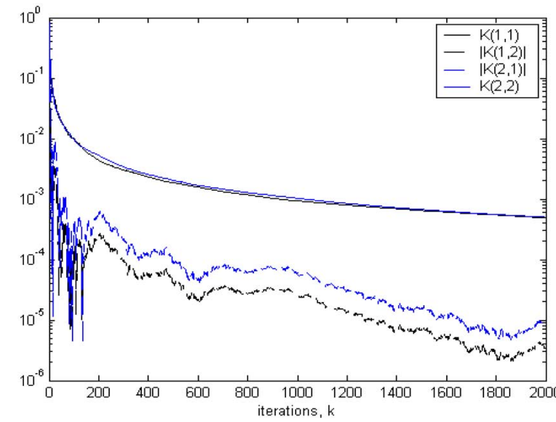

Fig. 4. Elements of the gain $K(k)$ for $\sigma = 2$.

- The values in Table I and Fig. 2 demonstrate the superiority of StNR over NR in presence of noisy measurements. For $\sigma \geq 2$, the errors associated with NR are $\gtrsim 100\%$. However, as σ decreases the gap in performance becomes smaller. Unlike NR, the long-term trend in errors associated with StNR keep on decreasing regardless the value of σ; e.g., see Fig. 3. The StNR values listed in Table I show that the errors at $k = 2000$ are roughly proportional to σ.
- Fig. 3 illustrates how well the error covariance matrix $P(k)$ can estimate the actual error variances. It is important to note that (14) is a set of nonlinear functions. Theorem 1 assumes a vector of linear functions (3). That is, more accurate estimates would be obtained if (14) were linear.
- Fig. 4 shows how the gain $K(k)$ decreases as the number of iterations increases. It is also important to note that $K(k)$ is diagonally dominate for all k.

IV. CONCLUSION

Proposition 1 of this letter showed that the traditional NR method cannot guarantee zero convergence in presence of additive measurement noise. Such convergence requires the consideration of adding a multiplicative iterative-varying gain matrix, which convergences to zero. This letter proposed a novel recursive algorithm providing optimal iterative-varying gain for linearized functions. It was analytically demonstrated that this proposed approach is capable of providing zero-convergence of the error covariance matrix for linear functions in presence of measurement noise. In addition, the proposed scheme was shown to overcome common drawbacks of NR method. An example has been presented illustrating the superiority of the proposed approach over the traditional NR method while taking into consideration noisy measurement functions.

ACKNOWLEDGMENT

The authors would like to thank Professor Samer S. Saab, the anonymous reviewers and associate editor for their constructive suggestions in improving this work.

REFERENCES

[1] J. Talvitie, M. Renfors, and E. S. Lohan, "Distance-based interpolation and extrapolation methods for RSS-based localization with indoor wireless signals," *IEEE Trans. Veh. Technol.*, vol. 64, no. 4, pp. 1340–1353, 2015.

[2] S. S. Saab and Z. Nakad, "A standalone RFID indoor positioning system using passive tags," *IEEE Trans. Ind. Electron.*, pp. 24–40, Jun. 2011.

[3] X. Zhang, J. Duan, Y. Fu, and A. Shi, "Theoretical accuracy analysis of indoor visible light communication positioning system based on received signal strength indicator," *J. Lightw. Technol.*, vol. 32, no. 21, pp. 4180–4186, Nov. 2014.

[4] S. Sakata and F. Ashida, "Ns-kriging based microstructural optimization applied to minimizing stochastic variation of homogenized elasticity of fiber reinforced composites," *Struct. Multidiscipl. Optim.*, vol. 38, pp. 443–453, 2009.

[5] W. Koch, *Tracking and Sensor Data Fusion Methodological Framework and Selected Applications*, Berlin, Germany: Springer, 2014.

[6] A. Abou Saleh, F. Alajaji, and W.-Y. Chan, "Compressed sensing with non-Gaussian noise and partial support information," *IEEE Signal Process. Lett.*, vol. 22, no. 10, pp. 1703–1707, 2015.

[7] W. Liu, S. Jia, L. Penglin, X. Chen, J. Yang, and Q. Wu, "An MRF-based depth upsampling: Upsample the depth map with its own property," *IEEE Signal Process. Lett.*, vol. 22, no. 10, pp. 1708–1712, 2015.

[8] B. Iooss, C. Lhuillier, and H. Jeanneau, "Numerical simulation of transit-time ultrasonic flowmeters: Uncertainties due to flow profile and fluid turbulence," *Ultrasonics*, vol. 40, no. 9, pp. 1009–1015, 2002.

[9] H. Robbins, S. Monro, "A stochastic approximation method," *Ann. Math. Statist.*, pp. 400–407, 1951.

[10] N. N. Schraudolph and T. Graepel, "Combining conjugate direction methods with stochastic approximation of gradients," in *Proc. 9th Int. Workshop on Artificial Intelligence and Statistics*, Key West, FL, USA, 2002, pp. 7–13.

[11] Z. Xu, L. Yingying, and X. Zhao, "Simulation-based optimization by new stochastic approximation algorithm," *Asia-Pac. J. Oper. Res.*, vol. 31, no. 4, 2014.

[12] B. Delyon and A. Juditsky, "Accelerated stochastic approximation," *SIAM J. Optim.*, vol. 3, pp. 868–881, 1993.

[13] J. C. Spall, *Introduction to Stochastic Search and Optimization*. Hoboken, NJ, USA: Wiley, 2003.

[14] J. C. Spall, "Adaptive stochastic approximation by the simultaneous perturbation method," *IEEE Trans. Automat. Contr.*, vol. 45, pp. 1839–1853, 2000.

[15] K. Hernandez and J. C. Spall, "Cyclic stochastic optimization with noisy function measurements," *Proc. Amer. Contr. Conf.*, Jun. 4–6, 2014, pp. 5204–5209.

[16] X. Hu, M. Bao, X. P. Zhang, and L. Guan, "Generalized iterated kalman filter and its performance evaluation," *IEEE Trans. Signal Process.*, vol. 63, no. 12, pp. 3204–3217, 2015.