How Important the Error Covariance in Simulated Kalman Filter?

Nor Hidayati Abdul Aziz, Zuwairie Ibrahim, Saifudin Razali
Faculty of Electrical and Electronics Engineering
Universiti Malaysia Pahang
26600 Pahang, Malaysia
pee14004@stdmail.ump.edu.my; {zuwairie; saifudin}@ump.edu.my

Taofiq Adeola Bakare, Nor Azlina Ab. Aziz
Faculty of Engineering and Technology
Multimedia University
75450 Melaka, Malaysia
1112700258@student.mmu.edu.my; azlina.aziz@mmu.edu.my

Abstract—The process of searching good parameter values is a non-trivial task for metaheuristic algorithms. When two algorithms are comparable in terms of speed and probability of convergence, the algorithm with less number of parameters is always preferred. This paper discussed the importance of the initial error covariance parameter, $\boldsymbol{P}(\mathbf{0})$, in Simulated Kalman Filter (SKF) with an intent to make SKF a parameter-less algorithm. To evaluate the importance of initial error covariance value in SKF, several values were selected and statistical analyses using nonparametric Friedman and Wilcoxon signed rank tests were carried out to see if different initial error covariance has any significant difference in the final outcome. The results prove that no matter what the initial error covariance is, SKF algorithm still managed to converge to near-optimal value without any significant degradation or improvement.

Keywords—SKF; Kalman; error covariance; metaheuristics; optimization

1. INTRODUCTION

Simulated Kalman Filter (SKF) was first introduced in 2015 by Ibrahim et al. as an optimizer for unimodal optimization problems [1]. Similar study on SKF actually has been carried out on all types of benchmark problems in CEC 2014 by Ibrahim et al. early this year, however, different algorithms were compared and the iteration numbers were increased [2]. Md. Yusof et al. have extended the SKF algorithm to deal with combinatorial optimization problems [3-5]. The discrete type SKF algorithm is then employed to solve Airport Gate Allocation Problems (AGAP) [6] and used as feature selection for peak detection of EEG signal [7]. Enhancement of the SKF algorithm has been carried out by Muhammad et al. by hybridizing the SKF algorithm with PSO algorithm [8]. All these studies suggest SKF is a good global optimizer.

Despite its good performance, SKF is not a parameter-free algorithm. Parameter tuning itself can be considered as an optimization problem. Some Evolutionary Algorithms (EA) has many parameters that are hard to tune. The challenges in EA is not only it requires a good initial parameter values, but some of the EA parameters have excessive sensitivities towards the overall performance. Genetic Algorithm (GA) for example [9] has many parameters to be set such as the probability of mutation and crossover, and the selection procedure. Particle Swarm Optimization (PSO) [10], despite being easy to understand, has 3 parameters to tune. Some classical algorithms, such as Tabu Search (TS) [11] and Simulated Annealing (SA) [12], has 2 and 1 parameters to tune respectively. Applications of such algorithms require some preliminary computation in order to tune the parameters before being applied to solve an optimization problem.

Covariance Matrix Adaptation Evolution Strategy (CMA-ES) [13] which is an evolutionary algorithm offers some default values to the users for its parameters. These values are applicable to any optimization problems. Self-tuning parameters like what are introduced to Differential Evolution (DE) [14] is an alternative solution. Ultimately parameter-free algorithms such as

Symbiotic Organisms Search (SOS) [15] is desirable. SOS is inspired by symbiotic relationship phenomena used by organisms to survive in an ecosystem. Therefore, this research is conducted as an attempt to reduce number of parameters to be tuned for SKF so that a parameter-free SKF can be achieved.

The next part of the paper is organized as follows: Section 2 talks about the error covariance in Kalman Filter and its relevance. Section 3 explains the experimental setup to test the significance of initial error covariance in SKF algorithm. Section 4 presents and discusses the results of the experiments, and Section 5 finally concludes the paper.

2. ERROR COVARIANCE IN KALMAN FILTER

Kalman Filter [16] is an estimation algorithm that makes use a series of noisy measurement observed over time to produce an estimate of unknown variables along with their uncertainties. The use of multiple measurements to form an estimate seems to give a better estimate rather than using a single measurement. Kalman Filter is a recursive algorithm of a two-step process. The first step is the prediction step where an estimate of the current state variables together with their uncertainties are estimated. Once the noisy measurement is observed, the estimation step takes place by improving the current estimates by averaging it with the newly observed measurement using a weighted average called Kalman gain. This results in a new state estimate that lies between the predicted and the measured state with a better level of uncertainty.

Covariance is a measure of estimated uncertainty. In Kalman Filter's process models, every measurement and calculations are estimated to a certain degree. The state in Kalman Filter at any particular time t is represented by two variables, the state estimate, $\hat{\boldsymbol{x}}_t$, and the error covariance matrix, P_t. Thus, the error covariance matrix, P_t, represents the estimated accuracy of the state estimate, $\hat{\boldsymbol{x}}_t$. Kalman Filter assumes the true state at any particular time t progressed from the previous time step $t-1$ following (1).

$$x_t = A_t x_{t-1} + B_t u_t + w_t \tag{1}$$

where A_t is the state transition matrix, B_t is the control input matrix, u_t is the control input vector, and w_t is the process noise vector. This process noise is assumed to be normally distributed with zero mean and a covariance Q_t. Measurements also comes with a certain amount of uncertainty. At a particular time t, the measurement, $\boldsymbol{z}_t$, of the true state, $\boldsymbol{x}_t$, is given by (2).

$$z_t = H_t x_t + v_t \tag{2}$$

where H_t is the measurement matrix, and v_t is the measurement noise vector. The measurement noise is assumed to be normally distributed white noise with zero mean and a covariance R_t. The initial state estimate and the noises are uncorrelated and independent of one another. These Gaussian-distribution assumptions just to ensure exact conditional probability estimate, and not a requirement in Kalman Filter.

The Kalman gain is a function of relative certainty between the current (predicted) state estimate and the estimated measurement given by (3).

$$K_t = P_{t|t-1} H_t^T \left(H_t P_{t|t-1} H_t^T + R_t \right)^{-1} \tag{3}$$

where $P_{t|t-1}$ is the predicted error covariance matrix. Kalman gain is a measure of trust between the two estimates. The high value of Kalman gain will make Kalman Filter follows measurements closely while low value of Kalman gain places more weight on the model predictions as given by (4).

$$\hat{x}_t = \hat{x}_{t|t-1} + K_t \left(z_t - H_t \hat{x}_{t|t-1} \right) \tag{4}$$

The current state estimate, $\hat{\boldsymbol{x}}_{t|t-1}$, and its corresponding covariance, $P_{t|t-1}$, are based on their previous values as shown in (5) and (6).

$$\hat{x}_{t|t-1} = A_t \hat{x}_{t-1} + B_t u_t \tag{5}$$

$$P_{t|t-1} = A_t P_{t-1} A_t^T + Q_t \tag{6}$$

Thus, the first iteration of Kalman Filter requires initialization values for these two variables. These initialization values are often not important because Kalman Filter will eventually converge the state estimate to the accurate value as iteration increases, thus reducing its covariance via the effect of Kalman gain as given by (7).

$$P_t = \left(I - K_t H_t \right) P_{t|t-1} \tag{7}$$

Poor selection of initial estimates is hypothesized to only extend the convergence time.

According to Kalman Filter principle, if the real process is accurately modelled, and the values of the initial state estimate, $\hat{\boldsymbol{x}}_0$, and the initial error covariance matrix, P_0, accurately represent the initial condition, then all the estimates should have a zero mean error. And if all the assumptions are followed and the Kalman Filter works optimally, the final error covariance will be a white noise and depicts the performance of the Kalman Filter.

3. EXPERIMENTS

The CEC 2014's benchmark suite is used in this paper. All 30 benchmark functions of CEC 2014 [17] are considered in the experiments. The SKF algorithm has 3 tuning parameters: the initial error covariance, $P(0)$, the process error, Q, and the measurement error, R. In this experiment, the effect of initial error covariance towards the algorithm's performance is investigated.

When SKF was first introduced, the initial error covariance, $P(0)$, was given the value of 1000, while the process error, Q, and the measurement error, R, were given the value of 0.5. To evaluate the effect of the initial error covariance, six experiments are conducted with different initial error covariance value while the other two parameters, the process and measurement error, are fixed at the same value of 0.5. Table 1 shows the settings of the tuning parameters. For all experiments, the common parameters such as the population size and maximum iterations, and other parameters of the experiments were given by Table 2. Note that all CEC 2014 benchmark functions are minimization problems and have the same search space of $[-100,100]$ for all dimensions. The search agents in SKF were initialized randomly within the search space for all benchmark functions.

Table 1: Tuning parameters.

	Initial error covariance, $P(0)$	**Process error, Q**	**Measurement error, R**
Experiment 1	0.01	0.5	0.5
Experiment 2	0.10	0.5	0.5
Experiment 3	1.00	0.5	0.5
Experiment 4	10.00	0.5	0.5
Experiment 5	100.00	0.5	0.5
Experiment 6	1000.00	0.5	0.5

Table 2: Experimental parameters.

Experimental parameters	**Values**
Number of agents	100
Number of iterations	10,000
Number of dimensions	50
Number of runs	50

To evaluate the significance of initial error covariance towards the performance of the SKF algorithm, the average performance from 50 independent runs of each experiment were statistically analysed using nonparametric Friedman and Wilconxon signed rank tests. The significance level of the tests was set to 0.05.

4. RESULTS AND DISCUSSION

The analytical results of the experiments for each benchmark function represented by their mean and standard deviation values are presented in Table 3. From the result in Table 3, we can see there is not much difference in SKF performance given a different initial error covariance value.

The mean values are then ranked using the Friedman test to see which initial error covariance value gives the best performance. In Friedman, the lower the rank indicates the better the performance Based on the rank, SKF gives the best performance when the initial error covariance, $P(0)$, is set to 1, and give the worst when is set to 1000. However, according to Friedman statistics, considering reduction performance distributed according to the chi-square distribution with 5 degrees of freedom, there are no significant difference exist on the SKF performance given a different value of P(0). This finding is consistent with Wilcoxon signed rank test finding. Wilcoxon signed rank test was used to compare 2 performances at one time. When SKF performance with initial error covariance of 1 and SKF performance with initial error covariance of 1000 are compared, the resulting p-value is more than 0.05. This means the performance difference between the two is not significant.

In both Friedman and Wilcoxon test, two performances are considered at par if the two-tail p-value is more or equal to the significance level of 0.05. The statistical results of Friedman and Wilcoxon tests are presented in Table 4. This result confirms our hypothesis that the value of the initial error covariance is not significant because of the effect of Kalman gain will reduce the initial error covariance, $P(0)$, eventually to a small constant value as the state estimate converges to a more accurate estimation.

Table 3: Analytical results of SKF with different $P(0)$.

Function Type	Function Name		$P(0) = 0.01$	$P(0) = 0.1$	$P(0) = 1$	$P(0) = 10$	$P(0) = 100$	$P(0) = 1000$
Unimodal	Rotated High Conditioned Elliptic	Mean	4.85E+06	4.34E+06	4.19E+06	4.47E+06	4.37E+06	4.54E+06
		Std Dev.	1.64E+06	1.54E+06	1.48E+06	2.03E+06	1.71E+06	1.33E+06
	Rotated Bent Cigar	Mean	6.19E+07	6.45E+07	1.27E+07	2.93E+07	8.89E+06	4.30E+07
		Std Dev.	1.92E+08	3.14E+08	3.84E+07	9.41E+07	3.16E+07	9.60E+07
	Rotated Discus	Mean	18347	16043	17821	15900	17970	16147
		Std Dev.	7255.8	7362.5	7193	6267.7	7453.8	7807.2
Simple Multimodal	Shifted and Rotated Rosenbrock's	Mean	523.79	529.92	531.28	532.47	524.37	521.37
		Std Dev.	38.242	36.733	43.207	39.659	40.878	37.614
	Shifted and Rotated Ackley's	Mean	520.01	520.01	520.01	520.01	520.01	520.01
		Std Dev.	0.020631	0.020129	0.014003	0.020577	0.013327	0.011198
	Shifted and Rotated Weierstrass	Mean	631.85	632.9	632.52	632.58	631.97	632.99
		Std Dev.	4.3205	4.2524	4.3402	4.4656	3.9574	4.7882
	Shifted and Rotated Griewank's	Mean	700.16	700.18	700.16	700.17	700.26	700.18
		Std Dev.	0.20458	0.25595	0.157	0.20203	0.35031	0.26646
	Shifted Rastrigin's	Mean	808.05	806.91	807.49	808.29	808.36	808.04
		Std Dev.	3.4977	3.6444	4.2356	5.2674	3.9612	4.8057
	Shifted and Rotated Rastrigin's	Mean	1059.8	1059.9	1061.6	1055.5	1064.2	1053.3
		Std Dev.	26.685	35.663	32.032	27.747	25.642	31.848
	Shifted Schwefel's	Mean	1343.8	1394.5	1388	1351.4	1407.7	1348.4
		Std Dev.	177.3	250.08	202.78	156.53	188.36	221.09
	Shifted and Rotated Schwefel's	Mean	6456.9	6245.6	6340.7	6272.6	6178.4	6257.8
		Std Dev.	696.52	818.76	782.7	747.41	809.77	934.97
	Shifted and Rotated Katsuura	Mean	1200.3	1200.3	1200.3	1200.3	1200.3	1200.3
		Std Dev.	0.091859	0.11644	0.080641	0.097888	0.075563	0.074247
	Shifted and Rotated HappyCat	Mean	1300.6	1300.6	1300.5	1300.6	1300.6	1300.5
		Std Dev.	0.080314	0.08644	0.075014	0.092012	0.090357	0.07784
	Shifted and Rotated HGBat	Mean	1400.3	1400.3	1400.3	1400.3	1400.3	1400.3
		Std Dev.	0.036195	0.038043	0.076739	0.035757	0.039213	0.035156
	Shifted and Rotated Expanded Griewank's plus Rosenbrock's	Mean	1553.7	1546.4	1557.7	1552.5	1551	1549.6
		Std Dev.	18.707	14.799	31.286	24.163	19.315	19.108
	Shifted and Rotated Expanded Scaffer's F6	Mean	1619	1619.1	1619.1	1619.1	1619	1619
		Std Dev.	0.76918	0.80714	0.83361	0.95534	0.82935	0.74572
Hybrid	Hybrid Function 1 (N=3)	Mean	8.32E+05	8.86E+05	8.99E+05	8.79E+05	8.91E+05	9.42E+05
		Std Dev.	4.21E+05	4.19E+05	4.75E+05	3.78E+05	5.20E+05	5.48E+05
	Hybrid Function 2 (N=3)	Mean	3.51E+06	7.43E+06	8.31E+06	7.14E+06	1.58E+07	4.04E+06
		Std Dev.	8.49E+06	1.52E+07	3.67E+07	3.15E+07	4.86E+07	1.31E+07
	Hybrid Function 3 (N=4)	Mean	1954	1948.9	1947.8	1947.8	1947.8	1950.6
		Std Dev.	30.295	29.356	30.835	30.879	31.172	28.084
	Hybrid Function 4 (N=4)	Mean	31025	35303	29036	30624	31454	34600
		Std Dev.	11312	14277	10675	14525	11940	17085
	Hybrid Function 5 (N=5)	Mean	1.24E+06	1.22E+06	1.28E+06	1.17E+06	1.14E+06	1.32E+06
		Std Dev.	5.29E+05	5.90E+05	6.21E+05	5.59E+05	4.30E+05	5.84E+05
	Hybrid Function 6 (N=5)	Mean	3429.7	3380.3	3386.4	3419	3386.4	3393.3
		Std Dev.	330.32	328.4	324.86	343.42	311.87	373.8
Composition	Composition Function 1 (N=5)	Mean	2645.4	2646.1	2645.7	2645.5	2645.8	2645.5
		Std Dev.	1.7839	2.9792	2.258	1.3595	2.4204	2.3682
	Composition Function 2 (N=3)	Mean	2664.4	2665.5	2664.3	2664.4	2664.3	2664.8
		Std Dev.	5.6832	6.0263	5.3134	6.0035	4.9565	6.1711
	Composition Function 3 (N=3)	Mean	2731.2	2730	2730.6	2731	2730.8	2730.5
		Std Dev.	4.0084	3.8477	4.5726	4.2195	3.6874	3.6144
	Composition Function 4 (N=5)	Mean	2786.4	2783.9	2774.4	2776.4	2788.4	2784.4
		Std Dev.	34.992	48.849	44.227	43.067	32.787	36.996
	Composition Function 5 (N=5)	Mean	3857.4	3877.2	3893.1	3902.3	3870.8	3868.2
		Std Dev.	97.747	104.88	127.98	111.07	104.45	97.063
	Composition Function 6 (N=5)	Mean	7275.7	7257.9	7080.7	6888.7	6928.8	7399.5
		Std Dev.	1277.6	940.1	1164.8	1028.6	1022.2	1025.5
	Composition Function 7 (N=3)	Mean	12873	17418	10668	7941.2	7109.4	17362
		Std Dev.	28612	76333	23407	10341	9596.1	51932
	Composition Function 8 (N=3)	Mean	19175	18710	18991	19076	19572	19644
		Std Dev.	2221.7	2995.2	3901.8	3154	3758.2	4538.2

Table 4: Statistical results of Friedman and Wilcoxon test.

P(0)	Friedman Test			Wilcoxon Signed Rank Test													
	Mean Rank	*p*-value	χ^2	R^+	R^-	*p*	R^+	R^-	*p*	R^+	R^-	*p*	R^+	R^-	*p*	R^+	R^-	*p*
0.01	3.6000	0.9622	1.0048	0.01 vs 0.1			0.01 vs 1			0.01 vs 10			0.01 vs 100			0.01 vs 1000		
0.1	3.6167			238	227	> 0.2	188	277	> 0.2	142	293	> 0.2	205	230	> 0.2	246	220	> 0.2
1	3.2833			0.1 vs 1			0.1 vs 10			0.1 vs 100			0.1 vs 1000			1 vs 10		
10	3.3333			250	215	> 0.2	149	286	> 0.2	234	231	> 0.2	254	211	> 0.2	208	227	> 0.2
100	3.5333			1 vs 100			1 vs 1000			10 vs 100			10 vs 1000			100 vs 1000		
1000	3.6333			235	230	> 0.2	285	180	> 0.2	212	223	> 0.2	284	181	> 0.2	280	185	> 0.2

4. CONCLUSION

The Simulated Kalman Filter algorithm, like most population-based metaheuristic optimization algorithm has common control parameters such as population size and maximum iterations. Besides these control parameters that affect the algorithm performance, the SKF algorithm, when it was first introduced, has 3 tuning parameters that need to be initialized before using the algorithm. Thus, this study is conducted to see if it is possible to reduce the number tuning parameters in SKF in order to make it a parameter-less algorithm. In the Kalman Filter framework, the value of initial error covariance $P(0)$ is not significant because eventually, its value will be reduced via the effect of Kalman gain to a very small value. This is the strength of Kalman Filter as a least-square-error estimator. By varying the initial error covariance value and fixing the other 2 tuning parameter values, this study proves that regardless any initializing value of the error covariance, the SKF performance still managed to converge to near optimal solution with no significant improvement or degradation. Hence, this study confirms that SKF can be applied to solve single objective optimization problems with only 2 tuning parameters, which are the process noise and the measurement noise.

ACKNOWLEDGMENT

This work was supported by Universiti Malaysia Pahang PGRS under Project ID GRS1503120 and Ministry of Higher Education FRGS under Project ID FRGS/1/2015/TK04/MMU/03/2.

REFERENCES

[1] Z. Ibrahim, N. H. Abdul Aziz, N. A. Ab. Aziz, S. Razali, M. I. Shapiai, S. W. Nawawi, and M. S. Mohamad, "A Kalman filter approach for solving unimodal optimization problems," ICIC Express Letters, vol. 9, no. 12, 2015, pp. 3415–3422.

[2] Z. Ibrahim, N. H. Abdul Aziz, N. A. Ab. Aziz, S. Razali, and M. S. Mohamad, "Simulated Kalman Filter: A novel estimation-based metaheuristic optimization algorithm", Advanced Science Letters, 2016, in press.

[3] Z. Md Yusof, Z. Ibrahim, I. Ibrahim, K. Z. Mohd Azmi, N. A. Ab. Aziz, N. H. Abdul Aziz, and M. S. Mohamed, "Angle modulated simulated Kalman filter algorithm for combinatorial optimization problems," ARPN Journal of Engineering and Applied Sciences, vol. 11, no. 7, 2015, pp.4854-4859.

[4] Z. Md Yusof, Z. Ibrahim, I. Ibrahim, K. Z. Mohd Azmi, N. A. Ab. Aziz, N. H. Abdul Aziz, and M. S. Mohamed, "Distance evaluated simulated Kalman filter algorithm for combinatorial optimization problems," ARPN Journal of Engineering and Applied Sciences, vol. 11, no. 7, 2015, pp.4904-4910.

[5] Z. Md Yusof, I. Ibrahim, S. Z. Satiman, Z. Ibrahim, N. H. Abdul Aziz, and N. A. Ab. Aziz, "BSKF: binary simulated Kalman filter," Third International Conference on Artificial Intelligence, Modelling and Simulation, 2015, pp. 77-81.

[6] Z. Md Yusof, S. Z. Satiman, K. Z. Mohd Azmi, B. Muhammad, S. Razali, Z. Ibrahim, Z. Aspar, and S. Ismail, "Solving airport gate allocation problem using Simulated Kalman Filter," In: International Conference on Knowledge Transfer 2015, 1-3 December 2015, Putrajaya Marriot Hotel, Malaysia.

[7] A. Adam, Z. Ibrahim, N. Mokhtar, M. I. Shapiai, M. Mubin, I. Saad, "Feature selection using angle modulated simulated Kalman filter for peak classification of EEG signals," SpringerPlus, Vol. 5, No. 1580, 2016.

[8] B. Muhammad, Z. Ibrahim, K. H. Ghazali, K. Z. Mohd Azmi, N. A. Ab. Aziz, N. H. Abdul Aziz, and M. S. Mohamed, "A new hybrid simulated Kalman filter and particle swarm optimization for continuous numerical optimization problems," ARPN Journal of Engineering and Applied Sciences, vol. 10, no. 22, 2015, pp.17171-17176.

[9] J. H. Holland, "Adaptation in natural and artificial systems," Ann Arbor: University of Michigan Press, 1975.

[10] J. Kennedy and R. Eberhart, "Particle swarm optimization," in: Proceedings of IEEE Int. Conf. on Neural Network, Perth, Australia, 1995, pp.1972—1978.

[11] F. Glover, "Future paths for integer programming and links to artificial intelligence," Computers and Operations Research, vol. 13, 1986, pp. 533–549.
[12] S. Kirkpatrick, C. Gelatt and M. Vecchi, "Optimization by simulated annealing," Science, vol. 220, 1983, pp. 671–680.
[13] N. Hansen, A. Ostermeier and A. Gawelczyk, "On the adaptation of arbitrary normal mutation distributions in evolution strategies: the generating set adaptation," in: Proceedings of the 6th International Conference on Genetic Algorithms, Morgan Kaufman Publishers Inc., San Francisco, CA, USA, 1995, pp. 57–64.
[14] R.M. Storn and K.V. Price, "Differential evolution – a simple and efficient heuristic for global optimization over continuous spaces," Journal of Global Optimization, vol. 11, 1997, pp. 341–359.
[15] M. Y. Cheng and D. Prayogo, "Symbiotic organisms search: a new metaheuristic optimization algorithm," Journal of Computers & Structures, vol. 139, 2014, pp. 98–112.
[16] R. E. Kalman, "A new approach to linear filtering and prediction problems," Trans. ASME-J. Basic Eng., vol. 82 (Series D), 1960, pp. 35–45.
[17] J. J. Liang, B. Y. Qu and P. N. Suganthan, "Problem definitions and evaluation criteria for the CEC 2014 special session and competition on single objective real-parameter numerical optimization," Tech. Rep. 201311, Computational Intelligence Laboratory, Zhengzhou University, 2013.