THE STOCHASTIC FILTERING PROBLEM: A BRIEF HISTORICAL ACCOUNT

BY DAN CRISAN

Abstract

Onwards from the mid-twentieth century, the stochastic filtering problem has caught the attention of thousands of mathematicians, engineers, statisticians, and computer scientists. Its applications span the whole spectrum of human endeavour, including satellite tracking, credit risk estimation, human genome analysis, and speech recognition. Stochastic filtering has engendered a surprising number of mathematical techniques for its treatment and has played an important role in the development of new research areas, including stochastic partial differential equations, stochastic geometry, rough paths theory, and Malliavin calculus. It also spearheaded research in areas of classical mathematics, such as Lie algebras, control theory, and information theory. The aim of this paper is to give a brief historical account of the subject concentrating on the continuous-time framework.

Keywords: Nonlinear filtering; Kalman–Bucy filter; Wiener filter; stochastic partial differential equation

2010 Mathematics Subject Classification: Primary 60G35; 93E11; 62M20
Secondary 35R60; 60H15; 97A30

The aim of nonlinear filtering is to estimate the current state of an evolving dynamical system, customarily modelled by a stochastic process, denoted by X and called the *signal process*. The signal process cannot be measured directly, but only via a related process Y, termed the *observation process*. The filtering problem consists in computing π_t, the conditional distribution of the signal X_t, at the current time t given the observation data accumulated up to that time. More precisely, π_t is the (random) probability measure which is measurable with respect to the observation filtration $\mathcal{Y}_t := \sigma(Y_s,\ s \in [0, t])$ so that

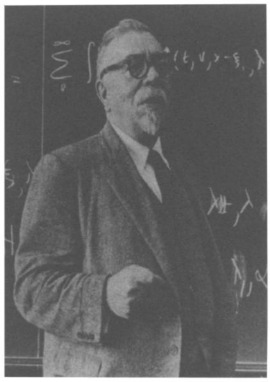
$$\mathbb{E}[\varphi(X_t) \mid \mathcal{Y}_t] = \int \varphi(x)\, \pi_t(\mathrm{d}x)$$

for all statistics φ for which both terms of the above identity make sense. Knowing π_t enables us, at least theoretically, to compute any inference of X_t given $\mathcal{Y}_t$, which is of interest, by integrating a suitable function φ with respect to π_t.

The origins of the filtering problem in discrete time can be traced back to the work of Kolmogorov and Krein [37, 38, 39, 40]. In the continuous-time case, N. Wiener was the first to discuss the optimal estimation of dynamical systems in the presence of noise. The Wiener filter consists of a signal X which is a stationary process and an associated measurement process $Y = X + V$, where V is some independent noise. The object is to use the values of Y to estimate X, where the estimate, $\widehat{X}_t$ say, is required to have the following three properties.

- *Causality*: X_t is to be estimated using Y_s for $s \le t$.

This account is an extended version of the historical notes included in [1, Section 1.3] and was presented at the 2014 Applied Probability Trust Jubilee Lectures in Sheffield.



A. N. Kolmogorov and N. Wiener

- *Optimality*: the estimate $\widehat{X}_t$ should minimise the mean square error $\mathbb{E}[(X - \widehat{X}_t)^2]$.
- *Online/concurrent estimation*: at any (arbitrary) time t, the estimate $\widehat{X}_t$ should be available.

The Wiener filter gives a linear, time-invariant causal estimate of the form

$$\widehat{X}_t = \int_{-\infty}^{t} h(t-s)Y(s)\,\mathrm{d}s,$$

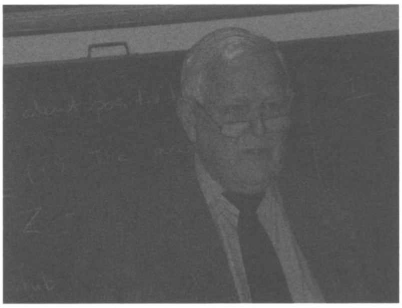
where $h(s)$ is called the transfer function. Wiener studied and solved this problem using the spectral theory of stationary processes. The results were included in a classified National Defense Research Council report issued in January/February 1942. The report, nicknamed 'The Yellow Peril' (according to Wiener [59] this was because of the yellow paper in which it was bound), was widely circulated among defence engineers. Subsequently declassified, it appeared as a book, [58], in 1949.

In the following quotation from [58], Wiener describes the historical relation between his work and that of Kolmogorov [37] on (almost) the same problem:

> ... The present investigation was initiated in the early winter of 1940 as an attempt to solve an engineering problem. At that time and until the last week of 1941, by which time the paper was substantially complete, the author was not aware of the results of Kolmogorov's work and scarcely aware of its existence; although Professor Feller of Brown University has mentioned Kolmogoroff's research in casual conversation at an earlier period. Mr I. E. Segal of the Princeton Graduate School brought Kolmogoroff's work to the author's attention at the Christmas meeting of the American Mathematical Society for 1941. Thus it would appear that the work of Kolmogoroff and that of the present writer represent two entirely separate attacks on the problem of time series, and that the parallelism between them may be attributed to the simple fact that the theory of stochastic processes had advanced to the point where the study of the prediction problem was the next thing on the agenda.

It is important to note that all consequent advances in the theory and practical implementation of stochastic filtering always adhered to the three precepts enumerated above: causality, optimality, and online estimation.

The next major development in stochastic filtering was the introduction of the *linear filter*. In this case the signal satisfies a stochastic differential equation with linear coefficients and Gaussian initial condition, and the observation equation satisfies an evolution equation with linear dependence on the signal. More precisely, let $X = (X^i)_{i=1}^d$ be the solution of the



© Mathematisches Forschungsinstitut Oberwolfach

R. Kalman

linear stochastic differential equation driven by a p-dimensional Brownian motion process $V = (V^j)_{j=1}^p$,

$$X_t = X_0 + \int_0^t (F_s X_s + f_s)\,\mathrm{d}s + \int_0^t \sigma_s\,\mathrm{d}V_s,$$

where, for any $s \geq 0$, F_s is a $d \times d$ matrix, σ_s is a $d \times p$ matrix, and f_s is a d-dimensional vector. The functions $s \mapsto F_s$, $s \mapsto \sigma_s$, and $s \mapsto f_s$ are measurable and locally bounded (that is, for every time t, the functions are bounded for $s \in [0, t]$). Assume that $X_0 \sim N(x_0, r_0)$ is independent of V. Next assume that W is a standard m-dimensional Brownian motion independent of X, and let Y be the process satisfying the evolution equation

$$Y_t = \int_0^t (H_s X_s + h_s)\,\mathrm{d}s + W_t, \tag{1}$$

where, for any $s \geq 0$, H_s is an $m \times d$ matrix and h_s is an m-dimensional vector (for further details, see, e.g. [1, Section 6.2]). The linear filter problem can be solved explicitly; in other words, π_t is given by a closed formula. The solution is a finite-dimensional one: π_t is Gaussian, namely, $\pi_t = N(\hat{x}_t, R_t)$, and, hence, completely determined by its mean $\hat{x}_t$ and its covariance matrix R_t. Moreover, it is quite easy to estimate the two parameters. The covariance matrix does not depend on Y and it satisfies the following deterministic Riccati equation:

$$\frac{\mathrm{d}R_t}{\mathrm{d}t} = \sigma_t^2 + 2F_t R_t - H_t^2 R_t^2, \qquad R_0 = r_0.$$

Hence, the covariance matrix can be computed in advance, before the filter is applied online. The mean satisfies the following linear stochastic differential equation driven by Y, whose solution can be easily computed:

$$\mathrm{d}\hat{x}_t = (F_t \hat{x}_t + f_t)\,\mathrm{d}t + R_t H_t (\mathrm{d}Y_t - (H_t \hat{x}_t + h_t)\,\mathrm{d}t), \qquad \hat{x}_0 = x_0.$$

As a result, the linear filter enjoyed widespread success in the 1960s; for example, it was used by NASA to get the Apollo missions off the ground and to the moon. R. S. Bucy and R. Kalman were the pioneers in this field. Kalman was the first to publish in a wide circulation journal. In [35] he solved the discrete-time version of the linear filter problem.

Bucy obtained similar results independently. Bucy and Kalman worked independently as the following quote taken from Bucy and Joseph [13] suggests:

> ... Later in the period 1957–1959, Bucy showed that the Follin example was just a special case of the general theory reported in this book (see Bucy [11]). Independently of this work Kalman in 1960, synthesizing his state space approach to linear systems and ideas in Doob [20], obtained a solution of the discrete-time optimal filter (see Kalman [35]).

> In April 1960, Bucy and Kalman became aware of each other's work and decided to collaborate on a paper dealing with continuous-time optimal filtering from the state space point of view and Bucy and Kalman [36] resulted.

Following the success of the linear filter, scientists started to explore different avenues. Firstly, they extended the application of the Kalman filter beyond the linear/Gaussian framework. The basis of this extension is the fact that, locally, all systems behave linearly. So, at least locally, one can apply the Kalman filter equation. This gave rise to a class of algorithms called the *Extended Kalman Filter*. These algorithms, most of which are empirical and without theoretical foundation, are still widely used today in a variety of applications.

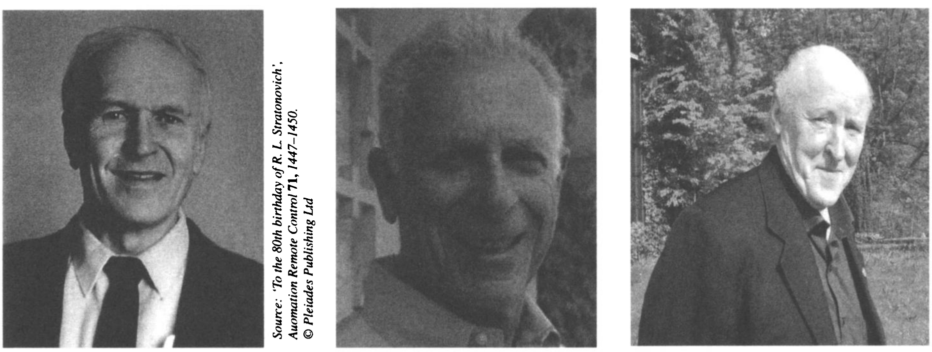
The development of the modern theory of nonlinear filtering started in the 1960s with the publications of R. L. Stratonovich, H. Kushner, and A. N. Shiryaev for diffusions, and W. M. Wonham for pure-jump Markov processes. Stratonovich's work in nonlinear filtering theory took place at the same time as the work of Bucy and Kalman. Stratonovich presented his first results in the theory of conditional Markov processes and the related optimal nonlinear filtering at the All-Union Conference on Statistical Radiophysics in Gorki (1958) and in a seminar [53]; they were published as [55].

Nevertheless, there was considerable unease about the methods used by Stratonovich to deduce the continuous-time filtering equation. His paper [55] appeared with an editorial footnote indicating that part of the exposition was not wholly convincing. Writing in *Mathematical Reviews*, Bharucha-Reid [4] indicated that he was inclined to agree with the editor's comment concerning the author's arguments in the continuous-time case.

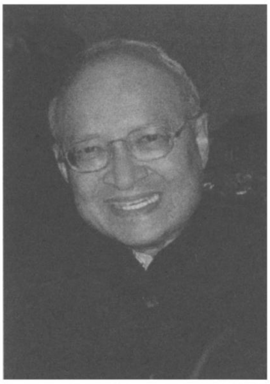
Part of the problem was that Stratonovich was using the stochastic integral, which today bears his name. Stratonovich himself mentioned this misunderstanding in [56, p. 42]. He also pointed out (*ibid.*, p. 227) that the linear filtering equations were published by him in [54].

On the other side of the Atlantic in the mid-1960s, Kushner [43, 44, 45] derived and analysed the evolution equation for the conditional distribution of the signal using Itô (and not Stratonovich) calculus. Separately, Shiryaev [51] provided the first rigorous derivation of the filtering equation in the case of a general observation process where the signal and observation noises may be correlated.

The equation satisfied by the conditional distribution of the signal was also obtained in various forms by other authors, namely, Bucy [12] and Wonham [61]. In 1968, Kailath [26] introduced the innovation approach to linear filtering. This new method for deducing the filtering equations was extended in the early 1970s by Frost and Kailath [22], and by Fujisaki, Kallianpur and Kunita [23]. Fujisaki, Kallianpur, and Kunita exploited a stochastic-integral representation



R. L. Stratonovich, H. Kushner, and A. N. Shiryaev

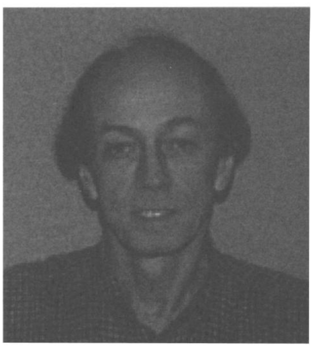


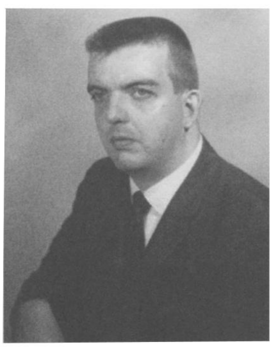
T. Kailath and H. Kunita

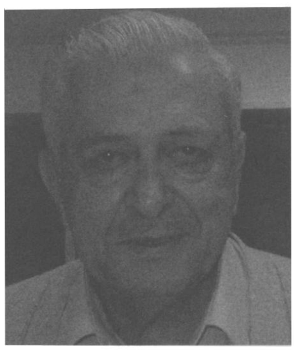
theorem in order to enable them to express conditional distributions as functionals of an 'innovation' martingale.

The filtering problem spearheaded the analysis of the equivalence between the observation filtration and that generated by the associated innovation martingale. Contributors to this direction include D. F. Allinger, V. E. Benes, J. M. C. Clark, A. Heunis, N. Krylov, and S. K. Mitter (see [15, Chapter 5] for details).

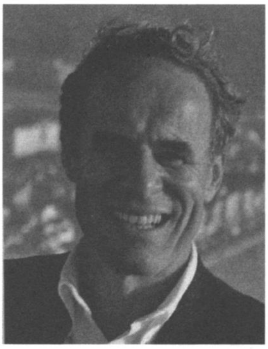
The filtering equation has inspired authors to introduce a rich variety of mathematical techniques to justify its structure. For example, although the filtering equation is nonlinear, it admits a linear version whose solution gives an unnormalized form of the condititional distribution of the signal. This linear evolution equation has been introduced by Duncan, Mortensen, and Zakai in [21, 46, 63]. Similarly, a lot of work has been done to establish the existence, uniqueness, and regularity of the solution of the filtering equation. A probabilistic approach, initially considered formally by Bucy [12], but developed in detail by Kallianpur and Striebel [33, 34], makes use of a functional form of Bayes formula for processes, now known as the Kallianpur–Striebel formula. This technique, which is based on a change of probability measure that makes, at each time, the future observation process independent of past processes, is effective for filtering problems in which the observation process is of the 'signal plus white noise' variety, where the signal is independent of the noise process, but less so for the 'correlated case', that is, for problems in which observed and unobserved components are coupled via a common noise process.

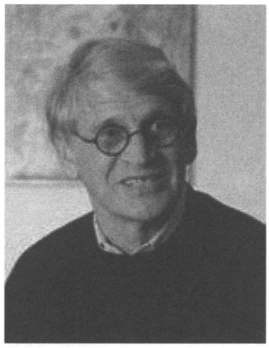






T. Duncan, R. Mortensen, and M. Zakai





E. Pardoux, N. Krylov, and B. Rozovsky

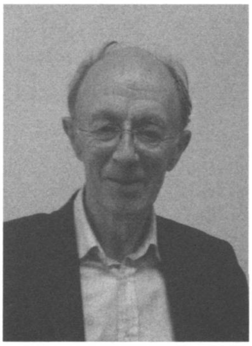
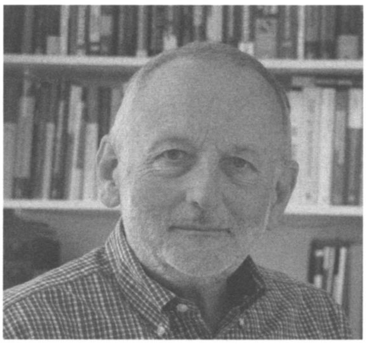
The stochastic partial differential equations associated with the filtering equations were rigorously analysed and extended in the late 1970s by Pardoux [47, 48, 49] and Krylov and Rozovsky [41, 42]. Pardoux adopted a functional analytic approach in analysing these stochastic partial differential equations, whereas Krylov and Rozovsky examined the filtering equations using methods inherited from classical partial differential equations theory.

There are other formulations of the filtering problem for which the signal process does not take the form of 'signal plus noise'. This occurs, for example, when the observation process is a counting or jump process with a jump rate modulated by the signal process. The study of the filtering problem for such processes was initiated by Snyder and Grigelionis [24, 52] in the early 1970s and developed rapidly with contributions from R. Boel, P. Brémaud, A. Segall, J. H. Van Schuppen, P. Varaiya, E. Wong [5, 6, 16, 50, 57, 60], and others. A comprehensive source of references in this direction can be found in [7].

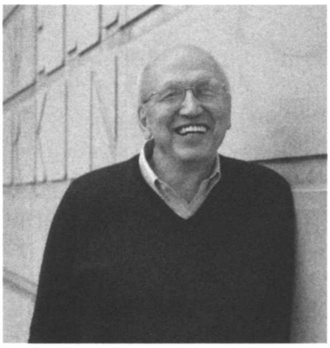
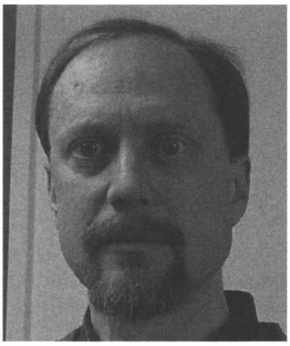
Indeed, there are many other approaches to analyse the filtering equations. We note the more recent work of B. I. Grigelionis and R. Mikulevicius on filtering for signal and observation processes with jumps [15, Chapter 4], and that of T. Kurtz and G. Nappo on the filtered martingale problem [15, Chapter 5]

Another important development in filtering theory was initiated by Clark [14] and continued by Davis [17, 19, 18]. In the late 1970s, Clark introduced the concept of *robust* or *pathwise* filtering, that is, that the filtering distribution admits a version which is a *continuous* function of the observation path. This in an important property from a computational point of view.

Given the success of the linear/Gaussian filter, scientists tried to find other classes of filtering problems where the solution is finite dimensional and/or has a closed form. Beneš [2] succeeded

J. M. C. Clark and M. H. A. Davis

R. W. Brockett and D. L. Ocone

in doing this by studying a class of filters with a linearly evolving observation process and for which the signal may have a nonlinear drift as long as it satisfies a certain (quite restrictive) condition, thenceforth known as the Beneš condition. The linear filter satisfies the Beneš condition.

Brockett [8, 9, 10] initiated a Lie algebraic approach to the filtering problem. From the linearized form of the filtering equation (the Zakai equation) one can deduce that ρ_t lies on a surface 'generated' by two differential operators. One is the infinitesimal generator of X, generally a second-order differential operator, and the other is a linear zero-order operator. From a Lie algebraic point of view, the Kalman filter and the Beneš filter are isomorphic, where the isomorphism is given by a state space transformation. Beneš continued his work in [3], where he found a larger class of exact filters for which the corresponding Lie algebra is no longer isomorphic with that associated with the Kalman–Bucy filter. Following Beneš, a number of authors further developed the Lie algebraic approach. Among the contributors in this direction are W. L. Chiou, J. Chen, G.-Q. Hu, C. W. Leung, D. L. Ocone, X. Wu, S. S.-T. Yau, and S. T. Yau. Others, such as F. Daum, S. J. Maybank, C. A. I. Schwartz, and B. W. Dickinson derived finite-dimensional filters by different techniques.

In contrast to these finite-dimensional filters, results have been discovered which prove that generically the filtering problem is infinite dimensional. Contributors in this direction include M. Chaleyat-Maurel, D. Michel, M. Hazewinkel, S. I. Marcus, H. J. Sussmann, and S. K. Mitter (many of the papers referred to here and earlier can be found in [25]). The general consensus now is that finite-dimensional filters are the exceptions and not the rule.

The work of Kallianpur has been influential in the field. The papers which contain the derivation of the Kallianpur–Striebel formula [33] and the derivation of the filtering equation [23] are of particular interest. Jointly with Karandikar, in a series of papers [27, 28, 29, 30, 31, 32], Kallianpur extended the theory of stochastic filtering to finitely additive measures in place of countably additive measures.

The area has expanded rapidly in the last twenty years. Among the topics developed in this period are stability of the solution of the filtering problem, the uniqueness and Feynman–Kac representations of the solutions of the filtering equations, Malliavin calculus applied to the qualitative analysis of the filtering distribution, and connections between filtering and information theory. In addition to the scientists already mentioned, contributions to the subject in this period were made by A. Bensoussan, A. Budhiraja, M. Chaleyat-Maurel, R. J. Elliott, B. Grigelionis, I. Gyöngy, M. Hazewinkel, A. Heunis, T. G. Kurtz, R. S. Liptser, D. Michel, R. Mikulevicius, N. J. Newton, J. Picard, W. J. Runggaldier, and O. Zeitouni.

Rapid progress continues to be made in both the theory and the application of stochastic filtering. In addition to work on the classical filtering problem, there is ongoing work on the analysis of the filtering problem for infinite-dimensional problems and problems where the Brownian motion noise is replaced by either 'coloured' noise, or fractional Brownian motion. Applications of stochastic filtering have been found in many areas of activity.

The development of numerical methods for the filtering problem has followed a parallel path with the theory being influenced by both theoretical advances and practical applications. We do not touch on this subject other than to mention that there is continuing work for developing both generic/universal numerical methods for solving the filtering problem and methods that are problem specific (see [1, Chapter 8] and [15, Parts VI–IX] for more detail).

Acknowledgement

The author would like to thank Martin Clark for his careful reading of the manuscript and for comments he made that led to several improvements of the historical account presented here.

References

[1] Bain, A. and Crisan, D. (2009). *Fundamentals of Stochastic Filtering* (Stoch. Modelling Appl. Prob. **60**). Springer, New York.

[2] Beneš, V. E. (1981). Exact finite-dimensional filters for certain diffusions with nonlinear drift. *Stochastics* **5,** 65–92.

[3] Beneš, V. E. (1985). New exact nonlinear filters with large Lie algebras. *Systems Control Lett.* **5,** 217–221.

[4] Bharucha-Reid, A. T. (1963). Review of 'Conditional Markov processes'. Available at http://www.ams.org/mathscinet/pdf/137157.pdf.

[5] Boel, R., Varaiya, P. and Wong, E. (1975). Martingales on jump processes. II. Applications. *SIAM J. Control* **13,** 1022–1061.

[6] Brémaud, P. (1976). La méthode des semi-martingales en filtrage quand l'observation est un processus ponctuel marque. In *Séminaire de Probabilités, X* (Lecture Notes Math. **511**), Springer, Berlin, pp. 1–18.

[7] Brémaud, P. and Jacod, J. (1977). Processus ponctuels et martingales: résultats récents sur la modélisation et le filtrage. *Adv. Appl. Prob.* **9,** 362–416.

[8] Brockett, R. W. (1981). Nonlinear systems and nonlinear estimation theory. In *Stochastic Systems: The Mathematics of Filtering and Identification and Applications* (Les Arcs, 1980; NATO Adv. Study Inst. Ser. C: Math. Phys. Sci. **78**), Reidel, Dordrecht, pp. 441–477.

[9] Brockett, R. W. (1984). Nonlinear control theory and differential geometry. In *Proceedings of the International Congress of Mathematicians* (Warsaw, 1983), eds Z. Ciesielski and C. Olech, PWN, Warsaw, pp. 1357–1368.

[10] Brockett, R. W. and Clark, J. M. C. (1980). The geometry of the conditional density equation. In *Analysis and Optimisation of Stochastic Systems* (Proc. Internat. Conf. Uni. Oxford, 1978), Academic Press, London, pp. 299–309.

[11] Bucy, R. S. (1959). Optimum finite time filters for a special nonstationary class of inputs. Tech. Rep. BBD 600, Applied Physics Laboratory, Johns Hopkins University.

[12] Bucy, R. S. (1965). Nonlinear filtering. *IEEE Trans. Automatic Control* **10,** 198.

[13] Bucy, R. S. and Joseph, P. D. (1987). *Filtering for Stochastic Processes with Applications to Guidance*, 2nd edn. Chelsea, New York.

[14] Clark, J. M. C. (1978). The design of robust approximations to the stochastic differential equations of nonlinear filtering. In *Communication Systems and Random Process Theory* (Proc. 2nd NATO Adv. Study Inst., Darlington, 1977; NATO Adv. Study Inst. Ser. E, Appl. Sci. **25**), ed. J. K. Skwirzynski, Sijthoff & Noordhoff, Alphen aan den Rijn, pp. 721–734.

[15] Crisan, D. and Rozovskii, B. (eds) (2011). *The Oxford Handbook of Nonlinear Filtering*. Oxford University Press.

[16] Davis, M. H. A. (1973). Non-linear filtering with point process observation. Res. Rep. 73/12, Department of Computing and Control, Imperial College London.

[17] Davis, M. H. A. (1980). On a multiplicative functional transformation arising in nonlinear filtering theory. *Z. Wahrscheinlichkeitsth.* **54,** 125–139.

[18] Davis, M. H. A. (1981). Pathwise nonlinear filtering. In *Stochastic Systems: The Mathematics of Filtering and Identification and Applications* (Les Arcs, 1980; NATO Adv. Study Inst. Ser. C: Math. Phys. Sci. **78**), eds M. Hazewinkel and J. C. Willems, Reidel, Dordrecht, pp. 505–528.

[19] DAVIS, M. H. A. (1982). A pathwise solution of the equations of nonlinear filtering. *Theory Prob. Appl.* **27,** 167–175.

[20] DOOB, J. L. (1953). *Stochastic Processes.* John Wiley, New York.

[21] DUNCAN, T. E. (1970). Likelihood functions for stochastic signals in white noise. *Inf. Control* **16,** 303–310.

[22] FROST, P. AND KAILATH, T. (1971). An innovations approach to least-squares estimation. Part III. Nonlinear estimation in white Gaussian noise. *IEEE Trans. Automatic Control* **16,** 217–226.

[23] FUJISAKI, M., KALLIANPUR, G. AND KUNITA, H. (1972). Stochastic differential equations for the non linear filtering problem. *Osaka J. Math.* **9,** 19–40.

[24] GRIGELIONIS, B. I. (1972). Stochastic equations for the nonlinear filtering of random processes. *Litovsk. Mat. Sb.* **12,** 37–51, 233 (in Russian).

[25] HAZEWINKEL, M. AND WILLEMS, J. C. (eds) (1981). *Stochastic Systems: The Mathematics of Filtering and Identification and Applications* (Les Arcs, 1980; NATO Adv. Study Inst. Ser. C: Math. Phys. Sci. **78**), Reidel, Dordrecht.

[26] KAILATH, T. (1968). An innovations approach to least-squares estimation. I. Linear filtering in additive white noise. *IEEE Trans. Automatic Control* **13,** 646–655.

[27] KALLIANPUR, G. (1985). White noise theory of filtering—Some robustness and consistency results. In *Stochastic Differential Systems* (Marseille-Luminy, 1984; Lecture Notes Control Inform. Sci. **69**), Springer, Berlin, pp. 217–223.

[28] KALLIANPUR, G. AND KARANDIKAR, R. L. (1984). The Markov property of the filter in the finitely additive white noise approach to nonlinear filtering. *Stochastics* **13,** 177–198.

[29] KALLIANPUR, G. AND KARANDIKAR, R. L. (1984). Measure-valued equations for the optimum filter in finitely additive nonlinear filtering theory. *Z. Wahrscheinlichkeitsth.* **66,** 1–17.

[30] KALLIANPUR, G. AND KARANDIKAR, R. L. (1985). A finitely additive white noise approach to nonlinear filtering: a brief survey. In *Multivariate Analysis VI* (Pittsburgh, PA, 1983), North-Holland, Amsterdam, pp. 335–344.

[31] KALLIANPUR, G. AND KARANDIKAR, R. L. (1985). White noise calculus and nonlinear filtering theory. *Ann. Prob.* **13,** 1033–1107.

[32] KALLIANPUR, G. AND KARANDIKAR, R. L. (1988). *White Noise Theory of Prediction, Filtering and Smoothing* (Stoch. Monogr. **3**). Gordon & Breach, New York.

[33] KALLIANPUR, G. AND STRIEBEL, C. (1968). Estimation of stochastic systems: arbitrary system process with additive white noise observation errors. *Ann. Math. Statist.* **39,** 785–801.

[34] KALLIANPUR, G. AND STRIEBEL, C. (1969). Stochastic differential equations occurring in the estimation of continuous parameter stochastic processes. *Theory Prob. Appl.* **14,** 567–594.

[35] KALMAN, R. E. (1960). A new approach to linear filtering and prediction problems. *J. Basic Eng.* **82,** 35–45.

[36] KALMAN, R. E. AND BUCY, R. S. (1961). New results in linear filtering and prediction theory. *J. Basic Eng.* **83,** 95–108.

[37] KOLMOGOROV, A. N. (1939). Sur l'interpolation et extrapolation des suites stationnaires. *C. R. Acad. Sci. Paris* **208,** 2043–2045.

[38] KOLMOGOROV, A. N. (1941). Interpolation and extrapolation of stationary random sequences. *Izv. Akad. Nauk. SSSR Ser. Math.* **5,** 3–14.

[39] KREIN, M. (1945). On a generalization of some investigations of G. Szegö, V. Smirnoff and A. Kolmogoroff. *C. R. (Doklady) Adad. Sci. URSS (N.S.)* **46,** 91–94.

[40] KREIN, M. (1945). On a problem of extrapolation of A. N. Kolmogoroff. *C. R. (Doklady) Akad. Sci. URSS (N.S.)* **46,** 306–309.

[41] KRYLOV, N. V. AND ROZOVSKIĬ, B. L. (1977). The Cauchy problem for linear stochastic partial differential equations. *Izv. Akad. Nauk SSSR Ser. Mat.* **41,** 1329–1347, 1448.

[42] KRYLOV, N. V. AND ROZOVSKIĬ, B. L. (1978). Conditional distributions of diffusion processes. *Izv. Akad. Nauk SSSR Ser. Mat.* **42,** 356–378, 470.

[43] KUSHNER, H. (1963). On the optimal control of systems with arbitrary distribution of initial position and noise corrupted observation. Tech. Rep. JA2123, Massachusetts Institute of Technology.

[44] KUSHNER, H. (1964). On the differential equations satisfied by conditional probability densities of Markov processes, with applications. *J. Soc. Indust. Appl. Math. Ser. A Control* **2,** 106–119.

[45] KUSHNER, H. J. (1967). Dynamical equations for optimal nonlinear filtering. *J. Differential Equat.* **3,** 179–190.

[46] MORTENSEN, R. E. (1966). Stochastic optimal control with noisy observations. *Internat. J. Control (1)* **4,** 455–464.

[47] PARDOUX, E. (1975). Equations aux dérivées partielles stochastiques non linéarires monotones. Doctoral Thesis, Université Paris XI, Orsay.

[48] PARDOUX, E. (1979). Stochastic partial differential equations and filtering of diffusion processes. *Stochastics* **3,** 127–167.

[49] PARDOUX, E. (1991). Filtrage non linéaire et équations aux dérivées partielles stochastiques associées. In *École d'Eté de Probabilités de Saint-Flour XIX—1989* (Lecture Notes Math. **1464**), Springer, Berlin, pp. 67–163.

[50] SEGALL, A., DAVIS, M. H. A. AND KAILATH, T. (1975). Non-linear filtering with counting observations. *IEEE Trans. Inf. Theory* **21,** 143–149.
[51] SHIRYAEV, A. N. (1966). On stochastic equations in the theory of conditional Markov processes. *Theory Prob. Appl.* **11,** 79–184.
[52] SNYDER, D. L. (1972). Filtering and detection for doubly stochastic Poisson processes. *IEEE Trans. Inf. Theory* **18,** 91–102.
[53] STRATONOVICH, R. L. (1959). On the theory of optimal non-linear filtration of random functions. *Theory Prob. Appl.* **4,** 223–225.
[54] STRATONOVICH, R. L. (1960). Application of the Markov processes theory to optimum filtration of signals. *Radio Eng. Electron. Phys.* **5,** 1–19.
[55] STRATONOVICH, R. L. (1960). Conditional Markov processes. *Theory Prob. Appl.* **5,** 156–178.
[56] STRATONOVICH, R. L. (1968). *Conditional Markov Processes and Their Application to the Theory of Optimal Control* (Modern Analytic Comput. Methods Sci. Math. **7**). Elsevier, New York.
[57] VAN SCHUPPEN, J. H. (1977). Filtering, prediction and smoothing for counting process observations, a martingale approach. *SIAM J. Appl. Math.* **32,** 552–570.
[58] WIENER, N. (1949). *Extrapolation, Interpolation, and Smoothing of Stationary Time Series: With Engineering Applications*. MIT Press, Cambridge, MA.
[59] WIENER, N. (1956). *I Am a Mathematician*. Doubleday, Garden City, NY.
[60] WONG, E. (1972). Martingale theory and applications to stochastic problems in dynamical systems. Res. Rep. 72/19, Department of Computing and Control, Imperial College London.
[61] WONHAM, W. M. (1965). Some applications of stochastic differential equations to optimal nonlinear filtering. *J. Soc. Indust. Appl. Math. Ser. A Control* **2,** 347–369.
[62] YAU, S.-T. AND YAU, S. S. T. (1998). Finite-dimensional filters with nonlinear drift. XI. Explicit solution of the generalized Kolmogorov equation in Brockett-Mitter program. *Adv. Math.* **140,** 156–189.
[63] ZAKAI, M. (1969). On the optimal filtering of diffusion processes. *Z. Wahrscheinlichkeitsth.* **11,** 230–243.

DAN CRISAN, *Imperial College London*

Department of Mathematics, Imperial College London, 180 Queen's Gate, London SW7 2AZ, UK.
Email address: d.crisan@imperial.ac.uk