DOI 10.2478/v10051-008-0020-8

Organization in Finance Prepared by Stochastic Differential Equations with Additive and Nonlinear Models and Continuous Optimization

Pakize Taylan[1,2], Gerhard-Wilhelm Weber[1,3]

[1]Middle East Technical University, Institute of Applied Mathematics, 06531 Ankara, Turkey, gweber@metu.edu.tr
[2]Dicle University, Department of Mathematics, 21280 Diyarbakır, Turkey, ptaylan@dicle.edu.tr
[3]University of Siegen, Faculty of Economics, Management and Law, 57076 Siegen, Germany

A central element in organization of financal means by a person, a company or societal group consists in the constitution, analysis and optimization of portfolios. This requests the time-depending modeling of processes. Likewise many processes in nature, technology and economy, financial processes suffer from stochastic fluctuations. Therefore, we consider stochastic differential equations (Kloeden, Platen and Schurz, 1994) since in reality, especially, in the financial sector, many processes are affected with noise. As a drawback, these equations are hard to represent by a computer and hard to resolve. In our paper, we express them in simplified manner of approximation by both a discretization and additive models based on splines. Our parameter estimation refers to the linearly involved spline coefficients as prepared in (Taylan and Weber, 2007) and the partially nonlinearly involved probabilistic parameters. We construct a penalized residual sum of square for this model and face occuring nonlinearities by Gauss-Newton's and Levenberg-Marquardt's method on determining the iteration step. We also investigate when the related minimization program can be written as a Tikhonov regularization problem (sometimes called ridge regression), and we treat it using continuous optimization techniques. In particular, we prepare access to the elegant framework of conic quadratic programming. These convex optimation problems are very well-structured, herewith resembling linear programs and, hence, permitting the use of interior point methods (Nesterov and Nemirovskii, 1993).

Key words: Stochastic Differential Equations, Regression, Statistical Learning, Parameter Estimation, Splines, Gauss-Newton Method, Levenberg-Marquardt's method, Smoothing, Stability, Penalty Methods, Tikhonov Regularization, Continuous Optimization, Conic Quadratic Programming.

1 Introduction

This paper is devoted to a modeling of financial processes which may serve as a basis of analysis and structural investigation. An important expression of this structure, the composition of its parts - its *organization* of financial assets - is called *portfolio* consisting of securities such as bonds, stocks, certificates, etc.. The organization of this portfolio requests pricing, hedging, optimization and optimal control. Those processes are on single assets and price processes, and on larger portfolios as well. The present study focusses on the first part of this modeling called regression, especially, parameter estimation.

Real-world data from the financial sector and science are often characterized by their great number and by a high variation. At the same time, the data need to become well understood and they have to serve as the basis of future prediction. Both the real situation and the practical requests are hard to balance (Hastie, Tibshirani and Friedman, 2001; Taylan and Weber, 2007; Taylan, Weber and Beck, 2007).

In fact, related mathematical modeling faces with non-differentiability and a high sensitivity of the model with respect to slightest perturbations of the data. Our paper analyzes this challenge by discussing and elaborating the corresponding parameter estimation problem by means of Tikhonov regularization, conic quadratic programming

and nonlinear regression methods. Herewith, we offer an alternative view and approach to stochastic differential equations (SDEs), and we invite to future research and practical applications. As a preparation, we firstly introduce into our methodology of statistical learning entitled *additive models*, which we will then exploit systematically. Indeed, we will apply them to SDEs, using modern methods of regularization and optimization. We shall address both the linear and the nonlinear case of parameter estimation. By this we develop and improve the results made in (Taylan and Weber, 2007).

2 Classical Additive Models

Regression models are very important in many applied areas, the *additive model* (Buja, Hastie and Tibshirani, 1989) is one of them. These models estimate an additive approximation of the multivariate regression function. For *N* observations on a response (or dependent) variable *Y*, denoted by $\boldsymbol{y} = (y_1, y_2, ..., y_n)^T$ measured at *N* design vector $\boldsymbol{x}_i = \left(x_{i1}, x_{i1}, ..., x_{im}\right)^T$, the additive model is defined by

$$Y = \beta_0 + \sum_{j=1}^{m} f_j\left(X_j\right) + \varepsilon, \tag{2.1}$$

where the errors ε are independent of the factors, X_j, $E(\varepsilon) = 0$ and $\mathrm{Var}(\varepsilon) = \sigma^2$. Here, the functions f_j are arbitrary unknown, univariate functions, they are mostly considered to be splines and we denote the estimates by $\hat{f}_j$. The standard convention consists in assuming at X_j that $E\left(f_j\left(X_j\right)\right) = 0$, since otherwise there will be a free constant in each of the functions (Hastie, Tibshirani and Friedman, 2001); all such constants are summarized by the intercept (bias) β_0.

1.1 Estimation Equations for Additive Model

Additive models have a strong motivation as a useful data analytic tool. Each function is estimated by an algorithm proposed by (Friedman and Stuetzle, 1981) and called *backfitting* (or *Gauss-Seidel*) *algorithm*. As our estimator for β_0, the mean of the response variable *Y* is used: $\hat{\beta}_0 = E(Y)$. This procedure depends on the partial residual against X_j:

$$r_j = Y - \beta_0 - \sum_{k \neq j} f_k\left(X_k\right), \tag{2.2}$$

and it consists of estimating each smooth function by holding all the other ones fixed . Then, $E(r_j \mid X_j) = f_j(X_j)$ which minimizes $E\left(Y - \beta_0 - \sum_{j=1}^{m} f_j\left(X_j\right)\right)^2$ (Friedman and Stuetzle, 1981; Hastie and Tibshirani, 1987).

3 Stochastic Differential Equations

3.1 Definition (Stochastic Differential Equations)

Many phenomena in nature, technology and economy are modelled by means of a deterministic differential equation with initial value $x_0 \in \mathbb{R}$:

$$\begin{cases} \dot{x}\ (:= dx/dt) = a(x,t), \\ x(0) = x_0. \end{cases}$$

But this type of modeling omits stochastic fluctuations and is not appropriate for, e.g., stock prices. To consider stochastic movements, *stochastic differential equation* (*SDE*) are used since they arise in modeling many phenomena, such as random dynamics in the physical, biological and social sciences, in engineering and economy. Solutions of these equations are often diffusion processes and, hence, they are connected to the subject of partial differential equations. We try to find a solution for these equations by an *additive* approximation (cf. Section 2), which is very famous in the statistical area, using spline functions.

Typically, a *stochastic differential equation*, equipped with an initial value, is given by

$$\begin{cases} \dot{X}(t) = a(X,t) + b(X,t)\delta_t & (t \in [0,\infty)), \\ X(0) = x_0, \end{cases} \tag{3.1}$$

where α is the deterministic part, $b\delta_t$ is the stochastic part, and δ_t denotes a generalized stochastic process (Kloeden, Platen and Schurz, 1994; Øksendal, 2003).

An example of a generalized stochastic processes is white noise. For a generalized stochastic processes, derivatives of any order can be defined. Suppose that W_t is a generalized version of a Wiener process which is used to model the motion of stock prices, which instantly responds to the numerous upcoming informations. A one-dimensional Wiener process (or a Brownian motion) is a time continuous process with the following properties.

1. $W_0 = 0$, with probability one.
2. $W_t \sim N(0,t)$ for all $t \in (0 \le t \le T)$, that is, for each t the random variable W_t is normally distributed with mean $E[W_t] = 0$ and variance $\mathrm{Var}[W_t] = E\left[W_t^2\right] = t$.
3. All increments $\Delta W_t =: W_{t+\Delta t} - W_t$ on nonoverlapping time intervals are independent. That is, the displacements $W_{t_2} - W_{t_1}$ and $W_{t_4} - W_{t_3}$ are independent for all $0 \le t_1 < t_2 \le t_3 < t_4$.

We note that a multi-dimensional Wiener processes can be similarly defined. Usually a Wiener process is differentiable almost nowhere. To obtain our approximate and, then, smoothened model, we treat W_t as if it was differentiable (a first approach which is widespread in literature). Then, white noise δ_t is defined as $\delta_t = \dot{W}_t = dW_t/dt$ and a Wiener process can be obtained by smoothing the white

noise. If we replace $\delta_t dt$ by dW_t in equation (3.1), then, this *stochastic differential equation* can be rewritten as

$$dX_t = a(X_t,t)dt + b(X_t,t)dW_t, \tag{3.2}$$

where $a(X_t,t)$ and $b(X_t,t)$ are drift and diffusion term, respectively, and X_t is a solution which we try to find based on the experimental data. Equation (3.2) is called ***Itô*** *SDE*. Here we want to simulate values of X_t, since we do not know its distribution. For this reason, we simulate a *discretized* version of the SDE.

3.2 Discretization of SDE

There are a number of discretization schemes available; we choose the ***Milstein scheme***. Then, we represent an approximation $\hat{X}_{t_j}$, in short: $\hat{X}_j$ $(j \in IN)$, of the process X_t by

$$\begin{aligned}\hat{X}_{j+1} &= \hat{X}_j + a(\hat{X}_j,t_j)(t_{j+1}-t_j) + b(\hat{X}_j,t_j)(W_{j+1}-W_j) \\ &\quad + \frac{1}{2}(b'b)(\hat{X}_j,t_j)\left[(W_{j+1}-W_j)^2 - (t_{j+1}-t_j)\right],\end{aligned} \tag{3.3}$$

where the prime "′" denotes the derivative with respect to *t*. Now, particularly referring to the finitely many sample (data) points $(\bar{X}_j,\bar{t}_j)$ $(j=1,2,...,N)$, we get

$$\begin{aligned}\dot{\bar{X}}_j &= a(\bar{X}_j,\bar{t}_j) + b(\bar{X}_j,\bar{t}_j)\frac{\Delta W_j}{\bar{h}_j} \\ &\quad + 1/2\,(b'b)(\bar{X}_j,\bar{t}_j)\left(\frac{(\Delta W_j)^2}{\bar{h}_j} - 1\right).\end{aligned} \tag{3.4}$$

Here, the value $\dot{\bar{X}}_j$ represents a difference quotient based on the *j* th experimental data $\bar{X}_j$ and on step lengths $\Delta\bar{t}_j = \bar{h}_j := \bar{t}_{j+1} - \bar{t}_j$ between neighbouring sampling times:

$$\dot{\bar{X}}_j := \begin{cases} \dfrac{\bar{X}_{j+1}-\bar{X}_j}{\bar{h}_j}, & \text{if } j=1,2,...,N-1, \\ \dfrac{\bar{X}_N-\bar{X}_{N-1}}{\bar{h}_N}, & \text{if } j=N. \end{cases}$$

The relations (3.4) cannot be expected to hold in an exact sense, since they include real data, but we satisfy them best in the *approximate* sense of least squares of errors. For the sake of convenience, we still write "=" instead of the approximation symbol "≈", and we shall study the least-squares estimation in Subsection 3.3.

Since $W_t \sim N(0,t)$, the increments ΔW_j are independent on non-overlapping intervals and moreover, $\mathrm{Var}(\Delta\bar{W}_j) = \Delta\bar{t}_j$, hence, the increments having normal distribution can be simulated with the help of standard normal distributed random numbers $\bar{Z}_j$. Herewith, we obtain a discrete model for a Wiener process:

$$\Delta\bar{W}_j = \bar{Z}_j\sqrt{\Delta\bar{t}_j},\ \bar{Z}_j \sim N(0,1). \tag{3.5}$$

If we use this value in our discretized equation, we obtain

$$\dot{\bar{X}}_j = a(\bar{X}_j,\bar{t}_j) + b(\bar{X}_j,\bar{t}_j)\frac{\bar{Z}_j}{\sqrt{\bar{h}_j}} + \frac{1}{2}(b'b)(\bar{X}_j,\bar{t}_j)\left(\bar{Z}_j^2-1\right). \tag{3.6}$$

For simplicity, we write equation (3.6) as

$$\dot{\bar{X}}_j = \bar{G}_j + \bar{H}_j c_j + (\bar{H}_j'\bar{H}_j)d_j, \tag{3.7}$$

where

$c_j := Z_j/\sqrt{\bar{h}_j}$, $d_j := 1/2\left(Z_j^2-1\right)$, $\bar{G}_j := a(\bar{X}_j,\bar{t}_j)$ and $\bar{H}_j := b(\bar{X}_j,\bar{t}_j)$.

To find the unknown values of $\bar{G}_j$ and $\bar{H}_j$, we consider the following optimization problem:

$$\min_y \sum_{j=1}^{N}\left(\dot{\bar{X}}_j - (\bar{G}_j + \bar{H}_j c_j + (\bar{H}_j'\bar{H}_j)d_j)\right)_2^2. \tag{3.8}$$

Here, *y* is a vector which comprises all the parameters in the Milstein model. We point out that also vector-valued processes could be studied, then referring to sums of terms in the Euclidean norm $\|\ \|_2^2$. Data from the stock market, but also from other sources of information or communcation, have a high variation.

Then, we must use a parameter estimation methods which will diminish this high variation and will give a smoother approximation to the data. ***Splines*** are more flexible and they allow us to avoid large oscillation observed for high-degree polynomial approximation. We recall that these functions can be described as linear combinations of basis splines and approximate the data $(\dot{\bar{X}}_j,\bar{t}_j)$ smoothly. Therefore, we approximate each function underlying the values $\bar{G}_j = a(\bar{X}_j,\bar{t}_j)$, $\bar{H}_j = b(\bar{X}_j,\bar{t}_j)$ and $\bar{F}_j = b'b(\bar{X}_j,\bar{t}_j)$ in an *additive* way established on basis splines. This treatment is very useful for the stability of the model in the presence of the many and highly varying data. Let us use basis splines for each function characterized by a separation of variables (coordinates); e.g., in equation (3.7):

$$\begin{aligned}\bar{G}_j &= a(\bar{X}_j,\bar{t}_j) = \alpha_0 + \sum_{p=1}^{2} f_p(\bar{U}_{j,p}) = \alpha_0 + \sum_{p=1}^{2}\sum_{l=1}^{d_p^g}\alpha_p^l\, B_p^l(\bar{U}_{j,p}), \\ \bar{H}_j c_j &= b(\bar{X}_j,\bar{t}_j)c_j = \beta_0 + \sum_{r=1}^{2} g_r(\bar{U}_{j,r}) = \beta_0 + \sum_{r=1}^{2}\sum_{m=1}^{d_r^h}\beta_r^m\, C_r^m(\bar{U}_{j,r}), \\ \bar{F}_j d_j &= b'b(\bar{X}_j,\bar{t}_j)d_j = \varphi_0 + \sum_{s=1}^{2} h_s(\bar{U}_{j,s}) = \varphi_0 + \sum_{s=1}^{2}\sum_{n=1}^{d_s^f}\varphi_s^n\, D_s^n(\bar{U}_{j,s}),\end{aligned} \tag{3.9}$$

where $\bar{U}_j = \left(\bar{U}_{j,1},\bar{U}_{j,2}\right) := \left(\bar{X}_j,\bar{t}_j\right)$. Let us give an example on how one can gain bases of splines. If we denote the *k*th order basis spline by $B_{\eta,k}$, a polynomial of degree *k* −1 with knots, say x_η, then a great benefit of using the base splines is provided by the following recursive algorithm (De Boor, 2001):

(3.10)

$$B_{\eta,1}(x)=\begin{cases}1, & \text{if } x_\eta \le x < x_{\eta+1}\\ 0, & \text{otherwise,}\end{cases}$$

$$B_{\eta,k}(x)=\frac{x-x_\eta}{x_{\eta+k-1}-x_\eta}B_{\eta,k-1}(x)+\frac{x_{\eta+k}-x}{x_{\eta+k}-x_{\eta+1}}B_{\eta+1,k-1}(x).$$

3.4 The Penalized Residual Sum of Squares Problem for SDE

We construct the *penalized residual sum of squares* for SDE in the following form:

(3.11)
$$PRSS(\theta,f,g,h):=\sum_{j=1}^{N}\left\{\dot{\bar{X}}_j-\left(\bar{G}_j+\bar{H}_jc_j+\bar{F}_jd_j\right)\right\}^2+\sum_{p=1}^{2}\lambda_p\int\left[f_p''(U_p)\right]^2dU_p$$
$$+\sum_{r=1}^{2}\mu_r\int\left[g_r''(U_r)\right]^2dU_r+\sum_{s=1}^{2}\varphi_s\int\left[h_s''(U_s)\right]^2dU_s.$$

Here, for convenience, we use the integral symbol "$\int$" as a dummy in the sense of $\int_{[a_\kappa,b_\kappa]}$, where $[a_\kappa,b_\kappa]$ $(\kappa=p,r,s)$ are appropriately large intervals where the integration takes place, respectively. Furthermore, $\lambda_p,\mu_r,\varphi_s\ge 0$ are *smoothing* (or *penalty*) *parameters*, they represent a tradeoff between first and second term. Large values of $\lambda_p,\mu_r,\varphi_s$ yield smoother curves, smaller values result in more fluctuation. If we use an additive form based on the basis splines for each function, then PRSS will become

(3.12)
$$\sum_{j=1}^{N}\left\{\dot{\bar{X}}_j-\left(\bar{G}_j+\bar{H}_jc_j+\bar{F}_jd_j\right)\right\}^2=$$
$$\sum_{j=1}^{N}\left\{\dot{\bar{X}}_j-\left(\alpha_0+\sum_{p=1}^{2}\sum_{l=1}^{d_p^g}\alpha_p^{\,l}\,B_p^l(\bar{U}_{j,p})+\beta_0+\sum_{r=1}^{2}\sum_{m=1}^{d_r^h}\beta_r^{\,m}\,C_r^m(\bar{U}_{j,r})+\varphi_0+\sum_{s=1}^{2}\sum_{n=1}^{d_s^f}\varphi_s^{\,n}\,D_s^n(\bar{U}_{j,s})\right)\right\}^2.$$

For simplicity, we introduce the following matrix notation:

(3.13)
$$\bar{G}_j+\bar{H}_jc_j+\bar{F}_jd_j=\alpha_0+\sum_{p=1}^{2}\sum_{l=1}^{d_p^g}\alpha_p^{\,l}\,B_p^l(\bar{U}_{j,p})+$$
$$+\beta_0+\sum_{r=1}^{2}\sum_{m=1}^{d_r^h}\beta_r^{\,m}\,C_r^m(\bar{U}_{j,r})+\varphi_0+\sum_{s=1}^{2}\sum_{n=1}^{d_s^f}\varphi_s^{\,n}\,D_s^n(\bar{U}_{j,s})$$
$$=\bar{A}_j\theta,$$

where

$$\bar{A}_j=\left(B_j\ \ C_j\ \ D_j\right),\ B_j=\left(1\ \ B_j^1\ \ B_j^2\right),\ C_j=\left(1\ \ C_j^1\ \ C_j^2\right),\ D_j=\left(1\ \ D_j^1\ \ D_j^2\right)$$
$$B_j^p=\left(B_p^1(\bar{U}_{j,p}),B_p^2(\bar{U}_{j,p}),...,B_p^{d_p^g}(\bar{U}_{j,p})\right)\ (p=1,2),\ C_j^r=\left(C_r^1(\bar{U}_{j,r}),C_r^2(\bar{U}_{j,r}),...,C_r^{d_r^h}(\bar{U}_{j,r})\right)(r=1,2),$$
$$D_j^s=\left(D_s^1(\bar{U}_{j,s}),B_s^2(\bar{U}_{j,s}),...,B_s^{d_s^f}(\bar{U}_{j,s})\right)(s=1,2)\ \text{ and}$$

$$\theta=\left(\alpha^T,\beta^T,\varphi^T\right)^T,\alpha=\left(\alpha_0,\alpha_1^T,\alpha_2^T\right)^T,\alpha_p=\left(\alpha_p^1,\alpha_p^2,...,\alpha_p^{d_p^g}\right)^T(p=1,2),\beta=\left(\beta_0,\beta_1^T,\beta_2^T\right)^T,$$
$$\beta_r=\left(\beta_r^1,\beta_r^2,...,\beta_r^{d_r^h}\right)(r=1,2),\ \varphi=\left(\varphi_0,\varphi_1^T,\varphi_2^T\right)^T,\varphi_s=\left(\varphi_s^1,\varphi_s^2,...,\varphi_s^{d_s^f}\right)^T(s=1,2).$$

Now, we can obtain the residual sum of squares as the squared length of the difference between $\dot{\bar{X}}$ and $\bar{A}\theta$, where $\bar{A}$ is matrix which contains the row vectors $\bar{A}_j$, and $\dot{\bar{X}}$ is our vector of difference quotients standing for the change rates of the experimental data:

(3.14)
$$\sum_{j=1}^{N}\left\{\dot{\bar{X}}_j-\bar{A}_j\theta\right\}^2=\left\|\dot{\bar{X}}-\bar{A}\theta\right\|_2^2,$$

where $\bar{A}=\left(\bar{A}_1^T,\bar{A}_2^T,...,\bar{A}_N^T\right)^T$, $\dot{\bar{X}}=\left(\dot{\bar{X}}_1,\dot{\bar{X}}_2,...,\dot{\bar{X}}_N\right)^T$.

Indeed, we get a discretized form of each integration term in the following way:

(3.15)
$$\int_a^b\left[f_p''(U_p)\right]^2dU_p\cong\sum_{j=1}^{N-1}\left[f_p''(U_{jp})\right]^2(U_{j+1,p}-U_{jp})$$
$$=\sum_{j=1}^{N-1}\left[\sum_{l=1}^{d_p^g}\alpha_p^lB_p^{l\prime\prime}(U_{jP})u_j\right]^2.$$

Using Riemann sums, we can discretize and represent each integration by the squared length of a vector, namely,

(3.16)
$$\int_a^b\left[f_p''(U_p)\right]^2dU_p\cong\sum_{j=1}^{N-1}\left[B_j^{p\prime\prime}u_j\alpha_p\right]^2=\left\|\bar{A}_P^B\alpha_p\right\|_2^2\quad(p=1,2),$$
$$\int_a^b\left[g_r''(U_r)\right]^2dU_r\cong\sum_{j=1}^{N-1}\left[C_j^{r\prime\prime}v_j\beta_r\right]^2=\left\|\bar{A}_r^C\beta_r\right\|_2^2\qquad(r=1,2),$$
$$\int_a^b\left[h_s''(\bar{U}_s)\right]^2dU_s\cong\sum_{j=1}^{N-1}\left[D_j^{s\prime\prime}w_j\varphi_s\right]^2D_j^s=\left\|\bar{A}_s^D\varphi_s\right\|_2^2\quad(s=1,2).$$

Here,

$$\bar{A}_p^B:=\left(B_1^{p\prime\prime T}u_1,B_2^{p\prime\prime T}u_2,...,B_{N-1}^{p\ \prime\prime T}u_{N-1}\right)^T,\ u_j:=\sqrt{U_{j+1,p}-U_{j,p}},$$
$$\bar{A}_r^C:=\left(C_1^{r\prime\prime T}v_1,C_2^{r\prime\prime T}v_2,...,C_{N-1}^{r\ \prime\prime T}v_{N-1}\right)^T,\ v_j:=\sqrt{U_{j+1,r}-U_{j,r}},$$
$$\bar{A}_s^D:=\left(D_1^{s\prime\prime T}w_1,D_2^{s\prime\prime T}w_2,...,D_{N-1}^{s\ \prime\prime T}w_{N-1}\right)^T,w_j:=\sqrt{U_{j+1,s}-U_{j,s}}\quad(j=1,2,...,N-1)$$

Using this discretized form in (3.17), PRSS looks as follows:

(3.17)
$$PRSS(\mathrm{q},f,g,h)=\left\|\dot{\bar{X}}-\bar{A}\theta\right\|_2^2+\sum_{p=1}^{2}\lambda_p\left\|\bar{A}_p^B\alpha_p\right\|_2^2+\sum_{r=1}^{2}\mu_r\left\|\bar{A}_r^C\beta_r\right\|_2^2+\sum_{s=1}^{2}\varphi_s\left\|\bar{A}_s^D\varphi_s\right\|_2^2$$

But, rather than a singleton, there is a finite sequence of the *tradeoff* or *penalty* parameters $\lambda=\left(\lambda_1,\lambda_2,\mu_1,\mu_2,\varphi_1,\varphi_2\right)^T$ such that this equation is not yet a *Tikhonov regularization problem* with a single such parameter. For this reason, let us make a uniform penalization by taking the same value $\lambda_p=\mu_r=\varphi_s=\lambda=\delta^2$ for each term. Then, our approximation of *PRSS* can be rearranged as

(3.18)
$$PRSS(\theta,f,g,h)=\left\|\dot{\bar{X}}-\bar{A}\theta\right\|_2^2+\delta^2\left\|\bar{L}\theta\right\|_2^2,$$

with the $(6(\acute{I}\ -1)\times m)$-matrix

$$\bar{L} := \begin{pmatrix} 0 & \bar{A}_1^B & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & \bar{A}_2^B & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \bar{A}_1^C & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & \bar{A}_2^C & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & \bar{A}_1^D & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \bar{A}_2^D \end{pmatrix}.$$

Herewith, based on the basis splines, we have identified the minimization of PRSS for some stochastic differential equation as an *Tikhonov regularization problem* (Aster, Borchers and Thurber, 2005):

$$\min_m \quad \|Gm-d\|_2^2 + \delta^2 \|Lm\|_2^2 \tag{3.19}$$

with penalty parameter $\lambda = \delta^2$. This regularization method is also known as *ridge regression*; it is very helpful for problems whose solution does not exist, or which is not unique or not stable under perturbations of the data. MATLAB Regularization Toolbox can be used for solution (Aster, Borchers and Thurber, 2005).

4 An Alternative Solution for Tikhonov Regularization Problem with Conic Quadratic Programming

4.1 Construction of the Conic Quadratic Programming Problem

We just mentioned that we can solve a Tikhonov regularization problem with MATLAB Regularization Toolbox. In addition, we shall explain how to treat our problem by using *continuous optimization* techniques which we suppose to become a complementary key technology and alternative to the concept of Tikhonov regularization. In particular, we apply the elegant framework of *conic quadratic programming* (*CQP*). Indeed, based on an appropriate, learning based choice of a bound *M*, we reformulate our Tikhonov regularization as the following optimization problem:

$$\begin{aligned} &\min_{\theta} \quad \left\| \bar{A}\theta - \dot{\bar{X}} \right\|_2^2, \\ &\text{subject to} \quad \left\| \bar{L}\theta \right\|_2^2 \le M. \end{aligned} \tag{4.1}$$

Here, the objective function in (4.1) is not linear but quadratic, however, the original objective function can be moved to the list of constraints, and we can write an equivalent problem as follows:

$$\begin{aligned} &\min_{t,\theta} \quad t, \\ &\text{subject to} \quad \left\| \bar{A}\theta - \dot{\bar{X}} \right\|_2^2 \le t^2, \quad t \ge 0, \\ &\qquad\qquad \left\| L\theta \right\|_2^2 \le M, \end{aligned} \tag{4.2}$$

or

$$\begin{aligned} &\min_{t,\theta} \quad t, \\ &\text{subject to} \quad \left\| \bar{A}\theta - \dot{\bar{X}} \right\|_2 \le t, \\ &\qquad\qquad \left\| L\theta \right\|_2 \le \sqrt{M}. \end{aligned} \tag{4.3}$$

Then, if we consider the form of a conic quadratic optimization problem (Nemirovski, 2002)

$$\min_x \ \boldsymbol{c}^T \boldsymbol{x}, \quad \text{subject to} \quad \|\boldsymbol{D}_i \boldsymbol{x} - \boldsymbol{d}_i\| \le \boldsymbol{p}_i^T \boldsymbol{x} - q_i \quad (i = 1,2,...,k), \tag{4.4}$$

we can see that our optimization problem for SDE is a conic quadratic program with $c = \begin{pmatrix} 1 & 0_m^T \end{pmatrix}^T$, $x = \begin{pmatrix} t & \theta^T \end{pmatrix}^T$, $D_1 = (0_N, \bar{A})$, $d_1 = \dot{\bar{X}}$,

$$p_1 = (1,0,...,0)^T, \quad q_1 = 0,$$

$$D_2 = (0_{6(N-1)}, L),\ d_2 = 0,\ p_1 = 0_{m+1}^T,\ q_1 = -\sqrt{M},$$

$$m = \sum_{p=1}^{2} d_p^g + \sum_{r=1}^{2} d_r^h + \sum_{s=1}^{2} d_r^h + 3.$$

In order to state the optimality conditions, we firstly reformulate our problem as

$$\min_{t,\theta} \quad t, \tag{4.5}$$

$$\text{such that} \quad \chi := \begin{pmatrix} 0_N & \bar{A} \\ 1 & 0_m^T \end{pmatrix} \begin{pmatrix} t \\ \theta \end{pmatrix} + \begin{pmatrix} -\dot{\bar{X}} \\ 0 \end{pmatrix},$$

$$\eta := \begin{pmatrix} 0_{6(N-1)} & \boldsymbol{L} \\ 0 & 0_m^T \end{pmatrix} \begin{pmatrix} t \\ \theta \end{pmatrix} + \begin{pmatrix} 0_{6(N-1)} \\ \sqrt{M} \end{pmatrix}.$$

Here, χ and η belong to L^{N+1} and $L^{6(N-1)+1}$, where L^{N+1} and $L^{6(N-1)+1}$ are the $(N+1)$- and $(6(N-1)+1)$-*dimensional ice-cream* (*or second-order Lorentz*) *cones*, defined by

$$L^v := \left\{ x = (x_1, x_2, ..., x_v)^T \in \mathsf{R}^v \mid x_v \ge \sqrt{x_1^2 + x_2^2 + ... + x_{v-1}^2} \right\} \ (v \ge 2).$$

Then, we can also write the *dual problem* to the latter problem as

$$\max \quad (\dot{\bar{X}}^T, 0)\kappa_1 + \left(0_{6(N-1)}^T, -\sqrt{M}\right)\kappa_2 \tag{4.6}$$

$$\text{such that} \quad \begin{pmatrix} 0_N^T & 1 \\ \bar{A}^T & 0_m \end{pmatrix} \kappa_1 + \begin{pmatrix} 0_{6(N-1)}^T & 0 \\ \boldsymbol{L}^T & 0_m \end{pmatrix} \kappa_2 = \begin{pmatrix} 1 \\ 0_m \end{pmatrix}$$

$$\kappa_1 \in L^{N+1}, \ \kappa_2 \in L^{6(N-1)+1}.$$

Morever, $(t, \theta, \chi, \eta, \kappa_1, \kappa_2)$ is the primal-dual optimal solution if the following constrains are provided in the corresponding ice-cream (second-order Lorentz) cones:

$$
\begin{aligned}
&\chi := \begin{pmatrix} 0_N & \bar{A} \\ 1 & 0_m^T \end{pmatrix} \begin{pmatrix} t \\ \theta \end{pmatrix} + \begin{pmatrix} -\dot{\bar{X}} \\ 0 \end{pmatrix}, \\
&\eta := \begin{pmatrix} 0_{6(N-1)} & \boldsymbol{L} \\ 0 & 0_m^T \end{pmatrix} \begin{pmatrix} t \\ \mathsf{q} \end{pmatrix} + \begin{pmatrix} 0_{6(N-1)} \\ \sqrt{M} \end{pmatrix}, \\
&\begin{pmatrix} 0_N^T & 1 \\ \bar{A}^T & 0_m \end{pmatrix} \kappa_1 + \begin{pmatrix} 0_{6(N-1)}^T & 0 \\ \boldsymbol{L}^T & 0_m \end{pmatrix} \kappa_2 = \begin{pmatrix} 1 \\ 0_m \end{pmatrix}, \\
&\kappa_1^T \chi = 0, \quad \kappa_2^T \eta = 0, \\
&\kappa_1 \in L^{N+1}, \quad \kappa_2 \in L^{6(N-1)+1}, \\
&\chi \in L^{N+1}, \quad \eta \in L^{6(N-1)+1}.
\end{aligned} \tag{4.7}
$$

4.2 On Solution Methods for Conic Quadratic Programming

For solving "well-structured" convex problems like conic quadratic problems, there are *interior point methods* (*IPM*s) which were firstly introduced by *Karmarkar* (1984). IPMs classically base on the interior points of the feasible set of the optimization problem; this set is assumed to be closed and convex. Then, an *interior penalty function* (*barrier*) $\boldsymbol{F}(\boldsymbol{x})$ is chosen, well defined (and smooth and strongly convex) in the interior of the feasible set. This function is "blowing up" as a sequence from the interior approaches a boundary point of the feasible set (Nesterov and Nemirovskii: 1993). Of great importance are *primal-dual IPMs* which refer to the pair of primal and dual variables.

The *canonical barrier function* for second–order (Lorentz) cones

$$\boldsymbol{L}^v := \{\, x = (x_1, x_2, ..., x_v)^T \in \mathsf{R}^n \mid x_v \geq \sqrt{x_1^2 + ... + x_{v-1}^2} \,\}$$

$(v \geq 2)$ is defined by $L_v(x) := -\ln(x_v^2 - x_1^2 - ... - x_{v-1}^2) = -\ln(x^T J_v x)$, where $J_v = \begin{pmatrix} -I_{v-1} & 0 \\ 0 & 1 \end{pmatrix}$. The *parameter* of this barrier is $\alpha(L_v) = 2$.

These algorithms have the advantage of employing the structure of the problem, of allowing better complexity bounds and exhibiting a much better practical performance.

5 On Nonlinear Dependence on Parameters and Their Estimation

Let return to equation (3.2) again, with two ways of generalization. (*i*) The model functions $a(\cdot)$ and $b(\cdot)$ may not only depend on the parameters which appear as coefficients in the linear combination with base splines, but also on really *probabilistic* (*stochastic*) parameters. (*ii*) Differently from the earlier linear dependence on the parameters, the dependence on the newly considered parameters may be *non*linear. In that case, we should use any nonlinear parameter estimation methods like, e.g., *Gauss-Newton's method* or *Levenberg-Marquardt's method* (Nash and Sofer, 1996).

Let us look at (*i*), for example, we consider following the stochastic differential equation:

$$\begin{cases} dX_t = \mathsf{m} X_t dt + \sigma X_t dW_t, \\ X(0) = x_0, \end{cases}$$

where $X_t = X(t)$ denotes the (random) price of a stock at time $t \geq 0$, and $\mu > 0$ and σ are parameters called the *drift* and *volatility* of the stock and x_0 is the starting price, respectively. Then, referring to the finitely many sample (data) points $(\bar{X}_\kappa, \bar{t}_\kappa)$ $(\mathsf{k} = 1, 2, ..., N)$ we get

$$\begin{aligned} \dot{\bar{X}}_\kappa &= \mathsf{m}\bar{X}_\kappa + \sigma \bar{X}_\kappa \frac{\Delta W_\kappa}{\bar{h}_\kappa} + \frac{1}{2}\sigma^2 (P'P)(\bar{t}_\kappa) \left(\frac{(\Delta W_\kappa)^2}{\bar{h}_\kappa} - 1 \right) \\ &= g(\bar{X}_\kappa, \mu, \sigma). \end{aligned}$$

To determine the unknown values μ, σ we consider following optimization problem:

(5.1)

$$\min f(\beta) := \sum_{\kappa=1}^{N} \left(\dot{\bar{X}}_\kappa - g(\bar{X}_\kappa, \mu, \sigma) \right)^2 =: \sum_{\kappa=1}^{N} f_\kappa^2(\beta) \quad \left(\text{or } \frac{1}{2}\sum_{\kappa=1}^{N} f_\kappa^2(\beta) \right)$$

Here, $\beta = (\mu, \sigma)^T$, $P(X) := X$, hence $P'(\bar{t}_\kappa) := 0$ (since P does not depend on t), and the objective function $f(\beta)$ of parameter estimation is defined linearly in auxiliary functions f_κ squared $(\kappa = 1, 2, ..., N)$. This problem representation holds true also if the quadratic term $(1/2)\sigma^2 (P'P)(\bar{t}_\kappa)\left((\Delta W_\kappa)^2 / \bar{h}_\kappa - 1\right)$ would not vanish and in many further examples where *(ii)* the parametric dependence may be *nonlinear* indeed.

Nonlinear parametric dependence can occur by the composition of stochastic processes. For example, in financial modelling of the dynamics of wealth from time t to $t + dt$ or maturity time T, V_t, may be given by

$$\begin{cases} dV_t = \left[\left(\theta_t^T (\mu - r\boldsymbol{e}) + r \right) V_t \right] dt - c_t dt + \theta_t^T \sigma V_t dW_t, \\ V_0 = v_0, \end{cases}$$

where θ_t is the fraction of wealth invested in the risky asset at time t and and c_t is the consumption at time t. We can easly identify both $a(t, V_t, c_t, \theta_t; r, \mu) := \left(\theta_t^T (\mu - r\boldsymbol{e}) + r \right) V_t - c_t$ and $b(t, V_t, \theta_t; \sigma) := \theta_t^T \sigma V_t$. Here, r is the short-term interest rate, μ denotes the vector of expected rates of return, $\boldsymbol{e}$ is the vector consisting of ones, σ stands the volatility matrix of the risky assets. The entire parameter $\beta := (r, \mu, \sigma)^T$ (arranged as a column vector) is assumed to be constant through time (Akume, 2007). Finally, W is a Wiener process with the property that dW_t is $N(0, dt)$ distributed. While the dependence of the right-hand side of the stochastic differential equation on β is linear, nonlinear parametric dependencies can occur via the insertion

of the processes c_t and θ_t in a and b, but also if r becomes a stochastic process r_t, e.g., in the following way. Namely, as a direct example of nonlinearity, the stochastic interest rate r_t for each $\tau \in \square$ may be given by

$$dr_t = \alpha \ (R - r_t)dt + \sigma_t \ r_t^{\tau} \ dW_t,$$

where σ_t and W_t are volatility and a Brownian motion, respectively Here, α is a positive constant, and the drift term $\alpha \ (R - r_t)$ is positive for $R > r_t$ and negative for $R < r_t$ (Seydel, 2003). We denote $a(t, r_t; R) := \alpha \ (R - r_t)$ and $b(t, r_t, \sigma_t; \tau) := \sigma_t r_t^{\tau}$. This process on the interest rate can be attached to a price or wealth process. By this interest rate processes and the composition of stochastic processes, further parameters such as (R, τ), can implicitly and in a partially nonlinear way enter the interest rate dynamics r_t and processes beyond of that dynamics.

In fact, the financial sector with the modeling and prediction of stock prices and interest rate are the most prominent application areas here. Moreover, mixed linear-nonlinear dependences on the parameters may be possible due to the linearly and the nonlinearly involved parameters of various kinds. This optimization problem (5.1) means a nonlinear least-squares estimation (or nonlinear regression). In the context of data fitting, each of the functions f_κ corresponds to a residual in our discrete approximation problem which may arise in a mathematical modelling or in an inverse problem. Let us represent basic ideas of nonlinear regression theory with the help of (Nash and Sofer, 1996).

Now, (5.1) can be represented in vector notation:

$$\min \ f(\beta) := \frac{1}{2} F^T(\beta) F(\beta), \tag{5.2}$$

where F is the vector-valued function $F(\beta) := (f_1(\beta), ..., f_N(\beta))^T \ (\beta \in \square^p)$ and where the factor $1/2$ serves for a more "optimal" normalization of the derivatives. In fact, by the chain rule we obtain

$$\nabla f(\beta) := \nabla F(\beta) F(\beta), \tag{5.3}$$

where $\nabla f(\beta)$ is an $(p \times N)$-matrix-valued function. By row-wise differentiation of $\nabla f(\beta)$ and using this gradient representation, we obtain the Hessian matrix of f:

$$\nabla^2 f(\beta) := \nabla F(\beta) \nabla F^T(\beta) + \sum_{\kappa=1}^{N} f_\kappa(\beta) \nabla^2 f_\kappa(\beta). \tag{5.4}$$

Let β^* be a solution of (5.1) and suppose $f(\beta^*) = 0$. Then, $f_\kappa(\beta^*) = 0 \ (\kappa = 1, 2, ..., N)$, i.e., all the residuals r_κ are vanishing and the model fits data without error. As a result, $F(\beta^*) = 0$ and, by (5.3), $\nabla f(\beta^*) = 0$, which just confirms our first-order necessary optimality condition. Furthermore, we can obtain the Hessian of f being

$$\nabla^2 f(\beta^*) := \nabla F(\beta^*) \nabla^T F(\beta^*),$$

which is a positive semi-definite matrix, just as we expected by our *second-order necessary optimality condition*. In case where $\nabla F(\beta^*)$ is a matrix of full rank, i.e., $\text{rank}(\nabla F(\beta^*)) = p$, then $\nabla^2 f(\beta^*)$ is positive definite, i.e., *second-order necessary optimality condition* is provided such that β^* is also a strict local minimizer.

From this basic idea, a number of specialized *nonlinear least-squares* methods come from. The simplest of this methods, called *Gauss-Newton* uses this approximative description in an indirect way. It make a replacement of the Hessian in the formula

$$\nabla^2 f(\beta) q = -\nabla f(\beta), \tag{5.5}$$

such that we have relation

$$\nabla F(\beta) \nabla^T F(\beta) q = -\nabla F(\beta) F(\beta), \tag{5.6}$$

where q is Gauss-Newton increment $q = \beta_1 - \beta_0$. If $F(\beta^*) \approx 0$ and $\text{rank}(\nabla F(\beta^*)) = p \ (\leq N)$, then, near to a solution β^*, Gauss-Newton behaves like Newton's method. However, we need not pay the computational cost of calculating second derivatives. Gauss-Newton's method sometimes behaves poor if there is one or a number of outliers, i.e., if the model does not fit the data well, or if $\text{rank}(\nabla F(\beta^*))$ is not of full rank p. In these cases, there is a poor approximation of the Hessian.

Many other nonlinear least-squares methods can be interpreted as using an approximation of the second additive form in the formula for the Hessian. i.e., of

$$\sum_{\kappa=1}^{N} f_\kappa (|\beta) \nabla^2 f_\kappa(\beta). \tag{5.7}$$

Levenberg-Marquardt's method uses the simplest of these approximation:

$$\sum_{\kappa=1}^{N} f_\kappa (|\beta) \nabla^2 f_\kappa(\beta) \approx | \ \mathrm{I}_p, \tag{5.8}$$

with some scalar $\lambda \geq 0$. This approximation yields the following linear system:

$$(\nabla F(\beta) \nabla^T F(\beta) + \lambda \ \mathrm{I}_p) q = -\nabla F(\beta) F(\beta). \tag{5.9}$$

We can often find Levenberg-Marquardt method implemented in the context of a trust-region strategy. There, q is obtained, e.g., by minimizing a quadratic model of the objective function with Gauss-Newton approximation of the Hessian:

(5.10)

$$\begin{cases} \min\limits_{q} \; Q(q) := f(\beta) + q^T \nabla F(\beta) F(\beta) + \frac{1}{2} q^T \nabla F(\beta) \nabla^T F(\beta) q \\ \text{subject to} \quad \|q\|_2 \le \Delta. \end{cases}$$

Here, λ is indirectly determined by picking a value of Δ. The scalar Δ can be chosen based on the effectiveness of the Gauss-Newton.

Levenberg-Marquardt method can be interpreted as a mixture between Gauss-Newton method (if $\lambda \approx 0$) and steepest-descent method (if λ is very large) (Aster, Borchers and Thurber, 2005; Nash and Sofer, 1996). An adaptive and sequential way of choosing λ and, by this, of the adjustment of mixture between the methods of Gauss-Newton and steepest-descent, is presented in (Nash and Sofer, 1996). We note that the term "λI_p" can also be regarded as a regularization term that shifts the eigenvalues of $\nabla F(\beta) \nabla^T F(\beta)$ away from 0.

Another way to solve the system (5.9) for given $\beta = \beta_k$, i.e., to find the (k+1)-st iterate $q = q_k$, constists in an application of least-squares estimation. If we denote (5.9) by $Gq = d$, where $G = \nabla F(\beta) \nabla^T F(\lambda) + \beta I_p$ and $d = -\nabla F(\beta) F(\beta)$, then we can study the regularized problem by adding to the squared residual norm $||Gq - d||_2^2$ a penalty or regularization term of the form $\delta^2 ||Lq||_2^2$, i.e.,

(5.11)

$$\min_{q} \quad \| \left(\nabla F(\beta) \nabla^T F(\beta) + \lambda I_N \right) q - \left(-\nabla F(\beta) F(\beta) \right) \|_2^2 + \delta^2 \| Lq \|_2^2,$$

where L may be the unit matrix, but it can also represent a discrete differentiation of first or second order. This regularization serves to diminish the complexity of the model. We recall (Aster, Borchers and Thurber, 2004) for closer explanation about this *Tikhonov regularization*. But instead of the penalization approach, we can again bound the regularization term $||Lq||_2^2$ by an inequality contraint. What is more, we can turn the optimization problem to a CQP problem in order to find the step q_k and, herewith, the next iterate $\beta_{k+1} := \beta_k + q_k$. By this conic quadratic modelling and solution technique we are back in the methodology that we presented in Section 4. Indeed, with a suitable and maybe adaptive choice of an upper bound M_1 (Işcanoğlu Çekiç, Weber and Taylan, 2007; Taylan and Weber, 2007; Taylan, Weber and Beck, 2007) we can write our problem as

(5.12)
$$\begin{aligned} &\min_{b} \quad \| \left(\nabla F(\beta) \nabla^T F(\beta) + \lambda I_N \right) q - \left(-\nabla F(\beta) F(\beta) \right) \|_2^2, \\ &\text{subject to} \quad \| Lq \|_2^2 \le M_1, \end{aligned}$$

or we can write an equivalent problem as follows:

$$\begin{aligned} &\min_{t,q} \quad t, \\ &\text{subject to} \quad \left\| \left(\nabla F(\beta) \nabla^T F(b) + \lambda I_p \right) q - \left(-\nabla F(\beta) F(\beta) \right) \|_2^2 \right\|_2^2 \le t^2, \quad t \ge 0, \\ &\qquad\qquad \| Lq \|_2^2 \le M_1. \end{aligned}$$

Then, if we consider the general problem form (Nemirovski, 2002)

$$\min_{x} \; \boldsymbol{c}^T \boldsymbol{x}, \quad \text{subject to} \quad \left\| \boldsymbol{D}_i \boldsymbol{x} - \boldsymbol{d}_i \right\| \le \boldsymbol{p}_i^T \boldsymbol{x} - q_i \quad (i = 1, 2, ..., k),$$

we can see that our optimization problem for determining step length q is a *conic quadratic program* with

$$c = \begin{pmatrix} 1 & 0_p^T \end{pmatrix}^T,$$

$$x = \begin{pmatrix} t & q^T \end{pmatrix}^T, \; D_1 = (0_p, \overline{A}), \; d_1 = -\nabla F(\beta) F(\beta),$$

$$p_1 = (1, 0, ..., 0)^T, \quad q_1 = 0,$$

$$D_2 = (0_p, L_{p \times p}), \; d_2 = 0_p, \; p_2 = 0_{p+1} \quad \text{and} \quad q_2 = -\sqrt{M_1},$$

6 Concluding Remarks

This paper gave a new contribution to problems related with SDEs using regression under an additive model or a nonlinear formulation, as a preparatory step on the way of *organizing* assets in terms of portfolios. We made modern methods of inverse problems and continuous optimization, especially, CQP and methods from nonlinear regression, become accessible and usable. Herewith, a bridge has been offered between statistical learning and data mining on the one hand, and the powerful tools prepared for well-structured convex optimization problems (Boyd and Vandenberghe, 2004) and Newton- and steepest-descent type regression methods (Nash and Sofer, 1996) on the other hand. We hope that future research, theoretical and applied achievements on this fruitful interface will be stimulated by our paper. The study on prediction of credit-default risk (Işcanoğlu Çekiç, Weber and Taylan, 2007) already showed the value of our generalized additive model approach. Indeed, further combined applications of our methods on real-word data from areas of finance, science and technology may be expected, where our contribution can be utilized.

References

Akume, D. A. (2007). *Risk Constrained Dynamic Portfolio Management*, Doctoral's Thesis, University of Yaounde I, Department of Mathematics.

Aster, R., Borchers, B. & Thurber, C. (2005). *Parameter Estimation and Inverse Problems*, Elsevier Academic Press, London.

Boyd, S. & Vandenberghe, L. (2004). *Convex Optimization*, Cambridge University Press, Cambridge.

Buja, A., Hastie, T. & Tibshirani, R. (1989). Linear smoothers and additive models, *The Ann. Stat.* **17**(2): 453-510.

De Boor, C. (2001). *Practical Guide to Splines*. Springer Verlag, New York.

Friedman, J. H. & Stuetzle, W. (1981). Projection pursuit regression, *J. Amer. Statist Assoc.* **76**: 817-823.

Hastie, T. & Tibshirani, R. (1987). Generalized additive models: some applications, *J. Amer. Statist. Assoc.* **82**(398): 371-386.

Hastie, T., Tibshirani, R. & Friedman, J.H. (2001). *The Element of Statistical Learning*, Springer Verlag, New York.

Işcanoğlu Çekiç, A., Weber G.-W. & Taylan, P. (2007). Predicting Default Probabilities with Generalized Additive Models for Emerging Markets, Invited Lecture, Graduate Summer School on New Advances in Statistics, METU, August 11-24, 2007. http://144.122.137.55/gweber/.

Kloeden, P. E, Platen, E. & Schurz, H. (1994). *Numerical Solution of SDE Through Computer Experiments*, Springer Verlag, New York.

Nash, G., and Sofer, A. (1996). *Linear and nonlinear programming*, McGraw-Hill, New York.

Nemirovski, A.(2002) Five lectures on modern convex optimization, Available from: http://iew3.technion.ac.il/Labs/Opt/opt/LN/Final.pdf (Accessed 5. April 2007).

Nesterov, Y. E. & Nemirovskii, A. S. (1994) *Interior Point Methods in Convex Programming*, SIAM Publications, Philadelphia.

Øksendal, B. K. (2003). *Stochastic Differential Equations: An Introduction with Applications*, Springer, Berlin.

Seydel, R. U. (2003). *Tools for Computational Finance*, Springer, Berlin.

Taylan, P. & Weber, G.-W. (2007). New approaches to regression in financial mathematics by additive models, *J. Comp. Techn.* **12**(2): 3-22.

Taylan, P. & Weber, G.-W. (2007) Approximation of stochastic differential equations by additive models using splines and conic programming, to appear in the proceedings of CASYS'07, Eighth International Conference on Computing Anticipatory Systems, Edited by Dubois, D. M. American Institute of Physics.

Taylan, P., Weber, G.-W. & Beck, A. (2007) New approaches to regression by generalized additive, models and continuous optimization for modern applications in finance, science and technology, *Optimization* 56(5–6): 675–698.

Pakize Taylan joined the Art and Science Faculty, Mathematics Department, of Dicle University in Diyarbakir, Turkey, in February 1990. She received an Msc. degree in 1993 and Ph.D. degree in 1999 from Dicle University in the field of mathematical statistics, and she earned her assistant professor degree in 2000. She tought courses in Turkish language, Probability and Statistics and Mathematical Statistics. She worked at Middle East Technical University, Ankara, at a Post-Doctoral position. Her research interests are liner regression, nonlinear regression, spline regression, optimization and Financial Mathematics. She has published journal articles and has presented her studies at national and international conferences. She still works as a head of Applied Mathematics at the Department of Mathematics, Art and Science Faculty, of her university.

Gerhard-Wilhelm Weber works at IAM of METU, Ankara, Turkey, in Department of Financial Mathematics and Department of Scientific Computing; he Assistant to the Director. Furthermore, he is Guest Professor at the Faculty of Economics, Management and Law of University of Siegen, Germany, and Collaborator at Center for Research on Optimization and Control (CEOC), University of Aveiro, Portugal. He received his Diploma and Doctorate in Mathematics and Economics at RWTH Aachen, and his Habilitation at TU Darmstadt. He held professorships by proxy at Institute of Mathematics, University of Cologne, and at Faculty of Mathematics, TU Chemnitz, Germany, then he worked at Cologne Bioinformatics Center; since 2003 he is at METU, Ankara. He chaired EUROPT where he is Past Chair now, he is Past Chair of EURO Working Group on OR in Computational Biology, Bioinformatics and Medicine, Vice Chair of EURO Working Group on OR for Development, Honoray Chair of EURO WG on Complex Societal Problems, and he represents German OR Society in Turkey. Prof. Weber got a number of awards, and he has been member in different research projects. His research interest are lying in the areas of Continuous Optimization, Financial Mathematics, Operations Research, Optimal Control, selected topics from Discrete Optimization, Dynamical Systems, Statistical Learning and Computational Statistics, Inverse Problems, Topology, Complexity Theory, Computational Biology and Bioinformatics, Environmental Protection, the sector of Development and Societal Complexity. Prof. Weber has published numerous articles, guest-edited 12 special issues, he has been in editorial boards of journals, wrote many referee and other reports. He co-organized about 50 scientific events (among of them also large ones of the EURO and INFORMS series) and is member in about 20 scientific organizations. Prof. Weber gave presentations all over the world, at scientific events and in seminars.