Interest rate models

Math 622

April 15, 2015

1 Introduction

So far in this class, we have studied various financial derivatives connected with a stock model. The stock is a typical example of a risky asset. On the other hand, we also have the bond, whose price is directly related to the interest rate, which in turn influences the price of the risk free asset: the money market account. In this Chapter, we will study various models of the short rate and the forward rate, which leads us to the price of the bond.

1.1 Money market account versus zero-coupon bond

It is clear that both the money market account and the zero-coupon bond prices are determined by the interest rate. But just how the two are similar and how are they different?

When the interest rate is a deterministic constant r, then the price at time t of a zero-coupon bond is $B(t,T) = e^{-r(T-t)}$. This is the same as the price of a money market account that has initial deposit at time 0 equals $e^{-rT} = B(0,T)$. On the other hand, it's also clear that the price at time t of a money market account with initial deposit K is Ke^{rt}, which is also the price of Ke^{rT} shares of zero-coupon with maturity T.

The situation is not the same when the interest rate is stochastic. First, observe that if the interest rate is an adapted process $R(t)$, then the value of the money market with an initial deposit K at time T is

$$M(T) = K\exp(\int_0^T R_u du),$$

and is random. Thus one cannot determine an initial deposit amount so that $M(T) = B(T,T) = 1$: the money market account cannot replicate a zero-coupon bond.

Similarly, the price at time 0 of a zero-coupon bond is

$$B(0,T) = \tilde{E}\Big(e^{-\int_0^T R_u du}\Big),$$

where $\tilde{E}$ is the expectation under a risk neutral measure. So without a prior assumption (or a model) of $R(t)$ under the risk neutral measure $\tilde{P}$, one cannot compute what $B(0,T)$ is (We say the bond price $B(0,T)$ is determined by the market).

An interesting question is by observing $B(t,T)$ for all $0 \le t \le T$, can one determine what $R(t)$ is? I believe for a fixed T, the answer is no. However, if we know the price of $B(t,T)$ for various maturity T, then we can deduce what $R(t)$ is, see the next section.

1.2 Various rates connected with bond

Since the payment of the bond with maturity T at time T is fixed, one can use the bonds (with various maturities, if necessary) to "lock-in" certain interest rates. Thus, zero-coupon bond prices are used as the standard for calculating interest rates. Throughout, it is assumed we deal with the market for risk free bonds and loans. The price at time $t \le T$ of a zero-coupon bond that pays \$1 at time T shall always be denoted by $B(t,T)$. Notice that $B(T,T) = 1$. There are various interest rates associated with B.

(i) Continuous compounding:

In this discussion all interest rates are quoted assuming continuous compounding. Consider an account which at time S has \$$L_S$ and at time $T > S$ has \$$L_T$, where S and T are measured in years. Then the interest rate r, per annum, continuously compounded, earned over $[S,T]$ is determined by the equation $L_S e^{r(T-S)} = L_T$, or

$$r = \frac{1}{T-S} \ln\left(\frac{L_T}{L_S}\right) \tag{1}$$

(ii) The spot rate

The *zero rate* for the period $[t,T]$, also called the *spot rate*, or, more precisely, the *continuously compounded spot rate for the period* $[t,T]$ is the function which gives the interest rate of a zero coupon bond over the interval $[t, T]$. That is, if we denote this rate by $R(t, T)$ then

$$1 = B(t, T) \exp\Big(R(t, T)(T - t)\Big).$$

From which it follows that

$$R(t, T) = \frac{1}{T - t} \ln\left(\frac{1}{B(t, T)}\right) = -\frac{\ln B(t, T)}{T - t} \tag{2}$$

Thus

$$B(t, T) = e^{-(T-t)R(t,T)}. \tag{3}$$

For a fixed t, a plot of $R(t, T)$ as a function of T is called a spot rate curve. It gives the (continuously compounded) interest rates available for risk free zero coupon bonds for all maturities starting from t. The notable fact about the spot rate curve is that it is not constant—normally it tends to be upward sloping. This phenomenon is called the term structure of interest rates.

(iii) The forward rate

Consider times $t < S < T$. Suppose at time t we would want to lock in certain spot rate for the time interval $[S, T]$. Let's call this rate $F(t, S, T)$. Then clearly this rate must be related with the bond price $B(t, S)$ and $B(t, T)$. So we should determine what $F(t, S, T)$ is and further inquire into whether we can indeed lock in this rate at time t by trading certain shares of the bonds with maturities at S and T.

To answer the first question, clearly what we want is if we invest 1 dollar at time S then we should receive $\exp\Big(F(t, S, T)(T - S\Big)$ at time T. Note that, 1 dollar at time S is equivalent to $B(t, S)$ at time t and $\exp\Big(F(t, S, T)(T - S)\Big)$ dollars at time T is equivalent to $B(t, T) \exp\Big(F(t, S, T)(T - S)\Big)$ at time t. These clearly should be equal if we want to lock in the rate $F(t, S, T)$ at time t. Therefore

$$B(t, S) = B(t, T) \exp\Big(F(t, S, T)(T - S)\Big),$$

or

$$F(t; S, T) = \frac{1}{T - S} \ln\left(\frac{B(t, S)}{B(t, T)}\right) = -\frac{\ln B(t, T) - \ln B(t, S)}{T - S}.$$

To answer the second question, first plug in $F(t,S,T) = -\frac{\ln B(t,T) - \ln B(t,S)}{T-S}$ into $B(t,T)\exp\Big(F(t,S,T)(T-S)\Big)$ and observe that

$$B(t,T)\exp\Big(F(t,S,T)(T-S)\Big) = B(t,T)\frac{B(t,S)}{B(t,T)}.$$

This suggests that we should hold $\frac{B(t,S)}{B(t,T)}$ shares of bond with maturity T at time t. To finance this position, we should sell 1 share of bond with maturity S at time t (since we expect to invest 1 dollar at time S). This turns out to be the right scheme because this costs nothing at time t: If at t we sell one zero-coupon bond maturing at S for $B(t,S)$ and with this money buy $B(t,S)/B(t,T)$ zero-coupon bonds maturing at T, the net value of this transaction for us is 0. At time S we pay out a dollar and at time T receive $B(t,S)/B(t,T)$. This is indeed equivalent to earning, at T, the amount $B(t,S)/B(t,T) = \exp\Big(F(t,S,T)(T-S)\Big)$ from a deposit of \$ 1 at S. The rate of interest earned by this transaction, $F(t,S,T)$ is called the *forward rate for* $[S,T]$ *contracted at* t.

Remark: $F(t,S,T)$ is known at time t by observing $B(t,S)$ and $B(t,T)$, that is $F(t,S,T) \in \mathcal{F}_t$, where $\mathcal{F}_t$ is the filtration generated by $B(t,S)$ and $B(t,T)$.

(iv) The instantaneous forward rate

The forward rate $F(t,S,T)$ has the formula

$$F(t;S,T) = -\frac{\ln B(t,T) - \ln B(t,S)}{T-S}.$$

If we let T goes to S, then the right hand side should go to $-\frac{\partial}{\partial T}\ln[B(t,S)]$, if the derivative exists. Indeed, if we assume $B(t,T)$ is differentible in T, then this is the case. This is not an unreasonable assumption since for a fixed t, one can believe that the bond price is a smooth function of different maturities. (On the other hand, for a fix maturity T, the bond price should *not* be a smooth function of t. It should be very irregular, indeed, in t, similar to behavior of the graph of a Brownian motion in t).

So we define the *instantaneous forward rate at t for investing at time T* as

$$f(t,T) = -\frac{\partial}{\partial T}\ln[B(t,T)]. \tag{4}$$

By integrating in T, it follows that

$$B(t,T) = \exp\{-\int_t^T f(t,u)\,du\} \tag{5}$$

For brevity, we refer to $f(t,T)$ as the forward rate function.

Remark: Again, note that here $f(t,T)$ is known at time t, that is $f(t,T) \in \mathcal{F}_t$ where $\mathcal{F}_t$ is the filtration generated by $B(t,T)$.

(v) The short rate

The *short rate* is the rate available at time t for the shortest period loans. Formally it is defined as

$$R(t) = f(t,t) \tag{6}$$

Remark: It seems reasonable to define $R(t) = f(t,t)$ (just from the understanding of what the forward rate is). First, it is reasonable to believe the spot rate $R(t,T)$ should converge to the short rate $R(t)$ when $T \to t$. Recall

$$R(t,T) = -\frac{\ln B(t,T)}{T-t} = \frac{\int_t^T f(t,u)du}{T-t}.$$

The RHS converges to $f(t,t)$ (by Lesbegue differentation theorem) as $T \to t$. So if we expect $R(t,T)$ to converge to $R(t)$, then it is reasonable to set $R(t) = f(t,t)$.

Second, from comparing the risk neutral pricing formula

$$B(t,T) = \tilde{E}\big(e^{-\int_t^T R(u)du}\big)$$

with the definition of the forward rate:

$$B(t,T) = e^{-\int_t^T f(t,u)du},$$

we should expect $R(t) = f(t,t)$ as well. Indeed, suppose $R(t) > f(t,t)$. Then if we suppose $R(u)$ and $f(t,u)$ are continuous functions of u (which is reasonable) then there must exist some $T > t$ so that $R(u) > f(t,u)$ for $u \in [t,T]$. But then we have

$$B(t,T) = e^{-\int_t^T f(t,u)du} > \tilde{E}\Big(e^{-\int_t^T R(u)du} \big| \mathcal{F}(t)\Big) = B(t,T),$$

which is a contradiction. So this cannot happen.

1.3 Remarks on modeling $B(t,T), R(t), f(t,T)$

At each t, the market presents us with the function $B(t,t+s)$, $s \geq 0$, or, equivalently, $f(t,t+s)$, $s \geq 0$, capturing, at each time t, the return on zero-coupon bonds of all maturities. As a function of s, this term structure of interest rates fluctuates as

t changes, and, we regard these fluctuations as random, because we cannot predict them exactly for future values of t. Developing good stochastic process models for the term structure of interest rates is a major area of mathematical finance. These models are used to analyze and price derivatives that depend on credit markets.

A few, loosely stated principles guide the construction of the basic models covered in this course. First, the models should be simple enough for fairly explicit calculation or at least easy simulation. Second, they should be rich enough that they can be calibrated to the market; that is, it should be possible to choose the parameters of the model so that its statistical behavior mimics reasonably well actual market performance. Of course, these two criteria push in opposite directions—the richer the model, the harder it is to analyze and simulate—and one must strike a good balance between them. A third important principle is that the model should not admit arbitrage.

From the previous section, all three processes $B(t,T), R(t), f(t,T)$ are objects of interest in modeling and we would like to obtain models for all three of them (for a fixed maturity T, as a process in t). It is also clear that if we get a model for one then the other two can be deduced out of it, via the relations:

$$\begin{aligned} B(t,T) &= e^{-\int_t^T f(t,u)du} \\ B(t,T) &= \tilde{E}\Big(e^{-\int_t^T R(u)du}\big|\mathcal{F}(t)\Big) \\ R(t) &= f(t,t). \end{aligned}$$

But there are subtle differences in which process we choose to model actually. First suppose we want to model $B(t,T)$ (which means we fix the maturity T and model $B(t,T)$ as a process in t). And let's say we go with the Geometric framework:

$$dB(t,T) = \alpha(t,T)B(t,T)dt + \sigma(t,T)B(t,T)dW_t,$$

under the physical measure P, where $B(0,T)$ is assumed known. The question is what should α and σ be? We recognize that they cannot be just any processes because we have the contraint:

$$B(T,T) = 1.$$

Indeed, unless for very trivial choices ($\sigma = 0$, α a constant) the terminal constraint cannot be easily satisfied. So modeling $B(t,T)$ as a function of t does not seem straightforward.

Note that the constraint $B(T,T) = 1$ is naturally built into the formula

$$B(t,T) \quad = \quad e^{-\int_t^T f(t,u)du}.$$

Thus, a good idea is to model the forward rate $f(t,T)$ and then derive $B(t,T)$ via the above formula. This is the HJM model and will be discussed in the next lecture.

The last approach is of course to model $R(t)$ and derive $B(t,T)$ via the relation

$$B(t,T) \quad = \quad \tilde{E}\Big(e^{-\int_t^T R(u)du} \big| \mathcal{F}(t)\Big).$$

The subtle point to note here is if we take this approach, then *necessarily we need to model* $R(t)$ *in under the risk neutral measure* $\tilde{P}$ *to obtain a price for* $B(t,T)$. The reason is this: if we model $R(t)$ in the physical probability P, say

$$dR(t) = \alpha(t)dt + \sigma(t)dW_t,$$

under P; then we do not have a market price of risk equation: there is one random source and there is no asset here. (Remember that we do not have the bond price dynamics under P under this approach, whose goal is to induce the bond price from the risk neutral pricing formula). Starting to model $R(t)$ under the risk neutral measure doesn't seem unreasonable; all it takes is to declare

$$dR(t) = \alpha(t)dt + \sigma(t)d\tilde{W}_t,$$

under $\tilde{P}$. But this may pose a problem for model calibration when we need to determine α, σ. The reason is we live in the physical world, i.e. we observe distribution under P. Nevertheless, modeling $R(t)$ under the risk neutral measure is an approach that many have taken and it leads to a connection with the forward rate HJM model, which can be modeled under the physical measure P. So in this lecture, we will discuss modeling the short rate $R(t)$ under the risk neutral measure and leave the connection with the forward rate for the next lecture.

The last question you may ask is, if we do not have the market price of risk equation, then who determines the risk neutral measure $\tilde{P}$? The answer is: the market does. I.e., it decides the bond price $B(t,T)$, which in turns imply what the risk neutral measure $\tilde{P}$ is (This is the answer given by Björk in his book: Arbitrage theory in continuous time). I'll leave it to you to ponder more about the meaning of this answer.

In both approaches (modeling $R(t)$ and $f(t,T)$), the models we present have an *ad hoc* flavor. It would seem more reasonable to build models supported by some theory of how the economy works. Such models would try to incorporate in a quantitative manner the economic factors and indicators that influence term structure. But we shall proceed innocent of all economic theory. We just look for models based on stochastic differential equations that we hope are rich enough to capture actual market behavior.

2 Multi-factor short rate models

A multi-factor model will consist of a vector-valued process

$$\mathbf{X}(t) = \begin{pmatrix} X_1(t) \\ \vdots \\ X_m(t) \end{pmatrix},$$

that solves a stochastic differential equation and is a Markov process under the risk-neutral measure, and a function $\Phi(x_1, \dots, x_m)$. The short rate is then **defined** by

$$R(t) = \Phi(X_1(t), \dots, X_m(t)).$$

The factors $X_i, i = 1, \cdots, m$ are meant to model economic factors that might influence the interest rate, such as GDP, import-export rate, inflation etc. A particularly simple choice for $\Phi(X_1(t), \dots, X_m(t))$ would be $R(t) = \delta_0(t) + \sum_{i=1}^m \delta_i(t) X_i(t)$. That is $R(t)$ is just a linear combination of the factors. One can then use linear regression to determine $\delta_i, i = 0, \cdots, m$.

Recall that the relation between the bond price and the short rate is via the equation

$$B(t,T) = \tilde{E}\left(e^{-\int_t^T R(u)du} \middle| \mathcal{F}(t)\right) \tag{7}$$

In the short-rate approach, being able to analyze the model comes down to calculating the conditional expectation in formula (7).

The key in evaluating this conditional expectation is an old idea that we have used over and over again in this course:

(i) Construct a model for the process $R(t)$ so that it is Markovian. Then (7) becomes

$$B(t,T) = \tilde{E}\left[\exp\{-\int_t^T \Phi(X_1(s), \dots, X_m(s))\, ds\} \,\middle|\, X_1(t), \dots, X_m(t)\right].$$

Thus we can find a function $c(t, x_1, x_2, ..., x_m)$ so that $c\big(t, X_1(t), X_2(t), ..., X_m(t)\big) = B(t,T)$.

(ii) Find a PDE that $c(t, x_1, x_2, ..., x_n)$ satisfies. If we can solve the PDE (via numerical procedure, for example) then we can recover the bond price $B(t,T)$ as described in (i).

2.1 Affine-yield model

Definition 2.1. *The short rate model is called an affine-yield model if it turns out that the zero-coupon bond price can be written as*

$$B(t,T) = \exp\{-C_1(t,T)X_1(t) - \cdots - C_m(t,T)X_m(t) - A(t,T)\},$$

for some functions $C_1(t,T), \ldots, C_m(t,T)$*, and* $A(t,T)$*.*

The Vasicek, CIR and Hull-White models discussed below are affine-yield models.

The case in which $m = 1$ is called a *single-factor model.* In such a model one takes $R(t)$ itself to be a Markov process; no auxiliary process $\mathbf{X}(t)$ is defined.

2.2 Connection with the forward rate

Affine yield models are particularly nice because it is easy to read off of an affine model a model for the instantaneous forward rate:

$$\begin{aligned} B(t,T) &= \exp\{-C_1(t,T)X_1(t) - \cdots - C_m(t,T)X_m(t) - A(t,T)\} \\ &= \exp(-\int_t^T f(t,u)du). \end{aligned}$$

Thus

$$-C_1(t,T)X_1(t) - \cdots - C_m(t,T)X_m(t) - A(t,T) = -\int_t^T f(t,u)du.$$

By differentiating both sides of the equation with respect to T, (assuming $C_i(t,T)$ and $A(t,T)$ are differentiable w.r.t T) we have a model for $f(t,T)$.

The obvious question is then how can we come up with candidates for affine yield models? We'll give an idea of how this is done in the two factor model in the section below. The generalization of this procedure for multi-factor model is straight forward.

3 Affine yield model in general

We illustrate the idea for two factor models under the risk neutral measure $\tilde{P}$,

$$\begin{aligned} dX_1(t) &= a_1(t, X_1(t), X_2(t))\,dt + b_{11}(t, X_1(t), X_2(t))\,d\widetilde{W}_1(t) + b_{12}(t, X_1(t), X_2(t))\,d\widetilde{W}_2(t) \\ dX_2(t) &= a_2(t, X_1(t), X_2(t))\,dt + b_{21}(t, X_1(t), X_2(t))\,d\widetilde{W}_1(t) + b_{22}(t, X_1(t), X_2(t))\,d\widetilde{W}_2(t) \end{aligned}$$

There is a lot of freedom in this general set-up, and we will quickly be more specific. But an easy first observation is that to obtain an affine yield model, we need $R(t)$ to be an affine function of X_1, X_2.

Thus we assume that $R(t) = \delta_0(t) + \delta_1(t)X_1(t) + \delta_2(t)X_2(t)$, where $\delta_i(t), i = 0, ...2$ are parameters of the model that would be determined by model calibration.

Because (X_1, X_2) is a Markov process, we know that $B(t,T) = g(t, X_1(t), X_2(t))$ for some function g.

To find an equation for g we start with the observation that

$$D(t)g(t, X_1(t), X_2(t)) = D(t)B(t,T)$$

is a martingale under the risk-neutral measure and that

$$g(T, X_1(T), X_2(T)) = B(T,T) = 1.$$

Apply Ito's formula,

$$\begin{aligned} d[D(t)B(t,T)] &= D(t)\mathcal{L}g(t, X_1(t), X_2(t))\,dt \\ &\quad + D(t)\mathcal{M}_1 g(t, X_1(t), X_2(t))\,d\widetilde{W}_1(t) + D(t)\mathcal{M}_2 g(t, X_1(t), X_2(t))\,d\widetilde{W}_2(t), \end{aligned}$$

where

$$\begin{aligned} \mathcal{L}g(t, x_1, x_2) &= -(\delta_0 + \delta_1(t)x_1 + \delta_2(t)x_2)g + g_t + a_1(t, x_1, x_2)g_{x_1} + a_2(t, x_1, x_2)g_{x_2} \\ &\quad + \frac{1}{2}[b_{11}^2 + b_{12}^2](t, x_1, x_2)g_{x_1x_1} + [b_{11}b_{21} + b_{12}b_{22}](t, x_1, x_2)g_{x_1x_2} \\ &\quad + \frac{1}{2}[b_{21}^2 + b_{22}^2](t, x_1, x_2)g_{x_2x_2} \end{aligned}$$

and

$$\begin{aligned} \mathcal{M}_1 g(t, X_1(t), X_2(t)) &= b_{11}(t, x_1, x_2)g_{x_1} + b_{21}(t, x_1, x_2)g_{x_2} \\ \mathcal{M}_2 g(t, X_1(t), X_2(t)) &= b_{12}(t, x_1, x_2)g_{x_1} + b_{22}(t, x_1, x_2)g_{x_2}. \end{aligned}$$

(In these expressions we have omitted writing the argument (t, x_1, x_2) of g and its partials.) In order that $D(t)g(t, X_1(t), X_2(t))$ be a martingale, g must be a solution of the parabolic pde,

$$\mathcal{L}g(t, x_1, x_2) = 0, \qquad t \leq T, \qquad g(T, x_1, x_2) = 1. \tag{8}$$

The following is the key observation:
If all the coefficients of the operator $\mathcal{L}$—that is

$$a_1(t, x_1, x_2), a_2(t, x_1, x_2), [b_{11}^2 + b_{12}^2](t, x_1, x_2), \text{ etc.}$$

are *affine functions*, i.e. functions of the form

$$\eta_0(t, T) + \eta_1(t, T)x_1 + \eta_2(t, T)x_2,$$

then (8) has a solution of the affine-yield form. That is, we can find $\alpha(t, T), c_1(t, T), c_2(t, T)$ such that

$$\begin{aligned} g(t, x_1, x_2) &= \exp\{-c_1(t, T)x_1 - c_2(t, T)x_2 - \alpha(t, T)\}, \quad \text{with} \\ c_1(T, T) &= c_2(T, T) = \alpha(T, T) = 0. \end{aligned}$$

Remark:

a. We require the $b_{11}^2 + b_{12}^2$ etc. to be affine, *not* b_{11} or b_{12} themselves. This explains the choice of the volatility in Vasicek model: constant in x_i and the choice of volatility in CIR model: $\sqrt{x_i}$ for b_{ii} and 0 for $b_{ij}, i \neq j$.

b. If $\eta_0(t, T), \eta_1(t, T), \eta_2(t, T)$ are constants, that is the coefficients in the affine form of $a_i(t, x_1, x_2$ etc. are constants, (which is the case for the standard Vasicek and CIR models) then we can check that $c_i(t, T)$ takes the form $c_i(T - t)$, similarly for $\alpha(t, T)$. That is their dependence on the two variables t, T is only on the difference $T - t$.

The conditions $c_1(T, T) = c_2(T, T) = \alpha(T, T) = 0$ imply that this function automatically satisfies the boundary condition $g(T, x_1, x_2) = B(T, T) = 1$.

Moreover, for this g, one can see by direct calculation that

$$\mathcal{L}g = D_1(t, T)x_1 g + D_2(t, T)x_2 g + D_3(t, T)g,$$

where $D_1(t,T), D_2(t,T), D_3(t,T)$ are defined **in terms of** $c_1(t,T), c_2(t,T), \alpha(t,T)$ and their first derivatives (in t).

Since $\mathcal{L}g(t, x_1, x_2) = 0$ we need the RHS of the above equation to be equal to 0 as well. This is accomplished by setting $D_i(t,T) = 0$, for all i.

So by setting $D_1(t,T) = 0$, $D_2(t,T) = 0$, $D_3(t,T) = 0$, *we obtained equations for determining* $c_1(t,T), c_2(t,T), \alpha(t,T)$ so that $\exp\{-c_1(t,T)x_1 - c_2(t,T)x_2 - \alpha(t,T)\}$ indeed solves (8).

If you examine the multi-factor CIR model or mixed models you will see that they are formulated exactly so that the coefficients of $\mathcal{L}g$ are affine. Following the derivation of the affine-yield expressions for these models in Shreve and doing Exercise 10.2 will help you understand this overall strategy.

4 Multi-factor Vasicek models

The single-factor Hull-White model is defined by a short rate which solves a linear differential equation of the sort, $dR(t) = (a(t) - b(t)R(t))\,dt + \sigma\, d\widetilde{W}(t)$, where $\widetilde{W}$ is a Brownian motion under the risk-neutral measure.

Multi-factor Vasicek models generalize the Hull-White model to the multi-factor case. In these models, $\mathbf{X}$ solves a linear stochastic differential equation with constant coefficients:

$$d\mathbf{X}(t) = A\mathbf{X}(t)\,dt + B\,d\widetilde{W}(t), \qquad \mathbf{X}(0) = \mathbf{X}_0 \tag{9}$$

where bX_0 is a given initial value, $\widetilde{W}$ is a multi-dimensional Brownian motion under the risk-neutral measure, and

$$R(t) = \delta_0 + \delta_1 X_1(t) + \cdots + \delta_m X_m(t).$$

Note that the choice of the liner coefficients $\delta_i, i = 0, \cdots, m$ in $R(t)$ are constants (not depending on t or ω). This is also a siginifcant simplifcation compared with the general multifactors model we proposed above.

Note also that no constant term $c\,dt$ enters (9). This causes no loss of generality if A is invertible, as we usually assume, because the constant term can be removed by an affine change of variables, as was shown in subsection (6.6) of this lecture.

Remark: The multi-factor Vasicek models have the advantage of having an explicit solution. More precisely, we can solve for an explicit formula for the factors $\mathbf{X}(t)$ -

see the dicussion in section (6). This in turn leads to *explicit computation* for $B(t,T)$, which we will discuss below.

On the other hand, if one wants a model so that the short rate $R(t)$ is non-negative, then the Vasicek model may not satisfy this condition (since $X(t)$ may be negative). Thus if imposing this non-negativity condition takes piority over the explicit solution, then one should go with the CIR model, discussed below. The CIR model does not have an explicit solution, but the interest rate $R(t)$ under the CIR model is guaranteed to be non-negative.

Example *Two-factor Vasicek; Canonical form.* The canonical form of the two-factor Vasicek is obtained by a linear change of variables to get an equivalent system with as few free parameters as possible. It is derived in Shreve assuming that W is a 2-dimensional Brownian motion and that A and B are invertible. The factors solve the system of Example 2 in Section (6.4) with $\sigma_1 = \sigma_2 = 1$:

$$d\begin{pmatrix} X_1(t) \\ X_2(t) \end{pmatrix} = \begin{pmatrix} -\lambda_1 & 0 \\ -\lambda_{21} & -\lambda_2 \end{pmatrix} \begin{pmatrix} X_1(t) \\ X_2(t) \end{pmatrix} dt + \begin{pmatrix} dW_1(t) \\ dW_2(t) \end{pmatrix} \tag{10}$$

The short rate is

$$R(t) = \delta_0 + \delta_1 X_1(t) + \delta_2 X_2(t) = \delta_0 + (\delta_1, \delta_2) \cdot \mathbf{X}(t). \qquad \diamond$$

4.1 Explicit formula for $B(t,T)$ under the Vasicek model

In this subsection, we will obtain an explicit formula for $B(t,T)$ under the Vasicek model. This requires some details about stochastic calculus in multi-dimensional model. These details will be presented in section (6).

Recall that the factors under Vasicek model have the following dynamics

$$d\mathbf{X}(t) = A\mathbf{X}(t)\,dt + B\,d\widetilde{W}(t), \qquad \mathbf{X}(0) = \mathbf{X}_0 \tag{11}$$

Recall that $\{\mathbf{X}(u);\, u \geq 0\}$ is a Gaussian process. Because of this, it can be shown that the conditional distribution of $\int_t^T R(u)\,du$ given $(X_1(t), \ldots, X_m(t))$ is Gaussian with a mean of the form

$$\tilde{E}[\int_t^T R(u)\,du \;\Big|\; X_1(t), \ldots, X_m(t)] = C_1(T-t)X_1(t) + \cdots + C_m(T-t)X_m(t) + (T-t)\delta_0,$$

where

$$(C_1(\tau), \ldots, C_m(\tau)) = \int_0^\tau (\delta_1, \cdots, \delta_m) e^{A\cdot u}\, du,$$

and with variance $\bar{A}(T-t)$, where

$$\bar{A}(\tau) = \int_0^\tau (C_1(u), \ldots, C_m(u)) BB^* (C_1(u), \ldots, C_m(u))^*\, du.$$

Here B^* denotes the transpose of B (the volatility in equation (11), don't confuse B with the bond price). $(C_1(u), \ldots, C_m(u))^*$ is the column vector that is the transpose of the row vector $(C_1(u), \ldots, C_m(u))$.)

Indeed, the explicit solution to equation (11) (given $\mathbf{X}(t)$ - see equation (23)) is

$$\mathbf{X}(u) = e^{A\cdot(u-t)}\mathbf{X}(t) + \int_t^u e^{A\cdot(u-s)} B\, d\tilde{W}(s).$$

Therefore

$$\begin{aligned}\int_t^T R(u)du &= \int_t^T \delta \cdot \mathbf{X}(u)du \\ &= \Big\{\int_t^T \delta \cdot e^{A\cdot(u-t)} du\Big\}\mathbf{X}(t) + \int_t^T \int_t^u \delta \cdot e^{A\cdot(u-s)} B\, d\tilde{W}(s)du.\end{aligned}$$

By switching the order of integration between $d\tilde{W}(s)$ and du, we have

$$\begin{aligned}\int_t^T \int_t^u \delta \cdot e^{A\cdot(u-s)} B\, d\tilde{W}(s)du &= \int_t^T \int_s^T \delta \cdot e^{A\cdot(u-s)} B du\, d\tilde{W}(s) \\ &= \int_t^T \int_0^{T-s} \delta \cdot e^{A\cdot u} B du\, d\tilde{W}(s) \\ &= \int_t^T \mathbf{C}(T-s) B\, d\tilde{W}(s),\end{aligned}$$

where

$$\mathbf{C}(\tau) := (C_1(\tau), \ldots, C_m(\tau)) = \int_0^\tau (\delta_1, \cdots, \delta_m) e^{A\cdot u}\, du,$$

is defined above.

We have

$$\int_t^T \mathbf{C}(T-s) B\, d\tilde{W}(s)$$

has Normal distribution with mean 0 and variance matrix

$$\int_t^T \mathbf{C}(T-s) BB^* \mathbf{C}^*(T-s) ds = \int_0^{T-t} \mathbf{C}(s) BB^* \mathbf{C}^*(s) ds.$$

Thus the distribution of $\int_t^T R(u)du = \int_t^T \delta \cdot \mathbf{X}(u)du$ is normal with mean

$$\Big\{\int_t^T \delta \cdot e^{A\cdot(u-t)}du\Big\}\mathbf{X}(t) \tag{12}$$

and variance

$$\int_t^T \mathbf{C}(T-s)BB^*\mathbf{C}^*(T-s)ds = \int_0^{T-t} \mathbf{C}(s)BB^*\mathbf{C}^*(s)ds. \tag{13}$$

It is trivial but helpful to keep in mind that the mean and variance here are real numbers, not vectors.

It follows from the formula $E[e^{\lambda Y}] = e^{\lambda\mu+\sigma^2\lambda^2/2}$ for the moment generating function of a normal random variable with mean μ and variance σ^2 that

$$\begin{aligned} B(t,T) &= \tilde{E}\left[\exp\{-\int_t^T R(u)\,du\} \mid X_1(t),\dots,X_m(t)\right] \\ &= \exp\{-C_1(T-t)X_1(t) - \cdots - C_m(T-t)X_m(t) - (T-t)\delta_0 + \frac{1}{2}\bar{A}(T-t)\}, \end{aligned}$$

where

$$\mathbf{C}(\tau) := (C_1(\tau),\dots,C_m(\tau)) = \int_0^\tau (\delta_1,\cdots,\delta_m)e^{A\cdot u}\,du$$

and

$$\bar{A}(\tau) = \int_0^\tau (C_1(u),\dots,C_m(u))BB^*(C_1(u),\dots,C_m(u))^*\,du.$$

are defined above. δ_0 is given from the model of $R(t)$.

Application of this formula to the canonical two-factor Vasicek model is carried out in this week's Assignment.

5 Cox-Ingersoll-Ross model

As mentioned above, the short rate $R(t)$ under Vasicek model can become negative. We consider instead the two-factor CIR model:

$$d\mathbf{X}(t) = \{\mu(t) - A(t)\mathbf{X}(t)\}\,dt + B(X_t)\,d\tilde{W}(t).$$

where

$$A(t) = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}, \qquad \mu(t) = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix},$$

$$B(X_t) = \begin{pmatrix} \sqrt{X_1(t)} & 0 \\ 0 & \sqrt{X_2(t)} \end{pmatrix}, \quad \tilde{W}(t) = \begin{pmatrix} \tilde{W}_1(t) \\ \tilde{W}_2(t) \end{pmatrix},$$

where $\mu_1, \mu_2, a_{11}, a_{22} > 0$ and $a_{12}, a_{21} \leq 0$.

The dynamics is constructed so that when $X_1(t) = 0$ then $\mu_1 - a_{12}X_2(t) \geq 0$ pushing X_1 above 0. Similarly for X_2. Thus one sees that $X_1(t), X_2(t)$ stays non-negative for all t if $X_1(0), X_2(0)$ are non-negative.

Thus by choosing

$$R(t) = \delta_0 + \delta_1 X_1(t) + \delta_2 X_2(t),$$

where we set $\delta_0 \geq 0, \delta_i > 0$ for all $i = 1, 2$ then $R(t)$ is non-negative as well.

There is no explicit solution for the CIR factor models (because of the $\sqrt{X_i(t)}$ term in the volatility). But also exactly because of this structure, and the independence of $\tilde{W}_1, \tilde{W}_2$ we also have an affine yield structure for the CIR model. In other words, the bond price $B(t,T)$ has the form

$$\begin{aligned} B(t,T) &= f(t, X_1(t), X_2(t)) \\ f(t, x_1, x_2) &= e^{-x_1 C_1(T-t) - x_2 C_2(T-t) - A(T-t)}. \end{aligned}$$

One can then set up a system of ODE equations for C_1, C_2, A in the fashion discussed in the Section (3) and solve for the bond price that way. See also Shreve's Section 10.2.2 for more details.

6 Linear Systems of Stochastic Differential Equations

6.1 The setting

This is a purely mathematical section. Linear systems of stochastic differential equations appear frequently in applied modeling and it is useful for the mathematical finance practitioner to know the basics about them.

By a linear system of stochastic equations we mean a system of the form

$$\begin{aligned}
dX_1(t) &= \{a_{11}(t)X_1(t) + a_{12}X_2(t) + \cdots + a_{1m}(t)X_m(t) + c_1(t)\}\ dt + \sum_{k=1}^{d} \sigma_{1k}(t)\, dW_k(t) \\
dX_2(t) &= \{a_{21}(t)X_1(t) + a_{22}X_2(t) + \cdots + a_{2m}(t)X_m(t) + c_2(t)\}\ dt + \sum_{k=1}^{d} \sigma_{2k}(t)\, dW_k(t) \\
\cdots &= \cdots\cdots\cdots\cdots\cdots \\
dX_m(t) &= \{a_{m1}(t)X_1(t) + a_{m2}X_2(t) + \cdots + a_{mm}(t)X_m(t) + c_m(t)\}\ dt + \sum_{k=1}^{d} \sigma_{mk}(t)\, dW_k(t)
\end{aligned}$$

It is much more efficient to write this in vector notation. Define

$$\mathbf{X} = \begin{pmatrix} X_1(t) \\ \vdots \\ X_m(t) \end{pmatrix}, \quad A(t) = \begin{pmatrix} a_{11}(t) & \cdots & a_{1m}(t) \\ \vdots & \vdots & \vdots \\ a_{m1}(t) & \cdots & a_{mm}(t). \end{pmatrix}, \quad c(t) = \begin{pmatrix} c_1(t) \\ \vdots \\ c_m(t) \end{pmatrix},$$

$$B(t) = \begin{pmatrix} \sigma_{11}(t) & \cdots & \sigma_{1d}(t) \\ \vdots & \vdots & \vdots \\ \sigma_{m1}(t) & \cdots & \sigma_{md}(t). \end{pmatrix}, \quad W(t) = \begin{pmatrix} W_1(t) \\ \vdots \\ W_d(t) \end{pmatrix}.$$

Then the system of equations becomes

$$d\mathbf{X}(t) = \{A(t)\mathbf{X}(t) + c(t)\}\, dt + B(t)\, dW(t). \tag{14}$$

Here $W(t)$ is a multi-dimensional Brownian motion, and $A(t)$, $c(t)$, and $B(t)$ are given functions of t. They could even be stochastic processes adapted to a filtration $\{\mathcal{F}(t);\, t \geq 0\}$ for W. In this lecture we shall present results only for the case in which $A(t) = A$, $B(t) = B$, and $c(t) = c$ are constant, deterministic matrices or vectors. This is the simplest, most often encountered case, and the theory for (14) is a fairly straightforward generalization from this case.

In equation (14) the components of $\mathbf{X}(t)$ do not appear in the '$dW(t)$' term. It is common to use the term "bilinear" for equations in which linear functions of the components of $\mathbf{X}(t)$ multiply $dW_i(t)$ terms. The Black-Scholes equation is bilinear in this sense.

Consider

$$d\mathbf{X}(t) = \{A\mathbf{X}(t) + c\}\, dt + B\, dW(t), \tag{15}$$

where A is an $m \times m$ matrix, B is an $m \times d$ matrix, c is an m-vector, and W is a d-dimensional Brownian motion. An explicit solution to this equation can be written down using the theory of ordinary linear systems of differential equations. This requires a bit of review.

6.2 The fundamental matrix for a linear system

Let I denote the $m \times m$ identity matrix. For a given $m \times m$ matrix A, define

$$e^{A \cdot t} = I + \sum_{k=1}^{\infty} A^k \frac{t^k}{k!}. \tag{16}$$

This is a matrix-valued, infinite series, and it can be proved that it converges for any A and any t, $-\infty < t < \infty$, and so $e^{A \cdot t}$ is well-defined.

Let $\frac{d}{dt} e^{A \cdot t}$ denote the matrix obtained by differentiating each entry of $e^{A \cdot t}$. Then it can be shown also that

$$\begin{aligned} \frac{d}{dt} e^{A \cdot t} &= \sum_{k=1}^{\infty} A^k \frac{d}{dt} \frac{t^k}{k!} = \sum_{k=1}^{\infty} A^k \frac{t^{k-1}}{(k-1)!} = A \cdot \left[I + \sum_{k=1}^{\infty} A^k \frac{t^k}{k!} \right] \\ &= A e^{A \cdot t}. \end{aligned} \tag{17}$$

Also, clearly, $e^{A \cdot 0} = I$. As a result, if $\mathbf{Z}_0$ is any m-vector, $Z(t) = e^{A \cdot t} \mathbf{Z}_0$ solves

$$\frac{d}{dt} \mathbf{Z}(t) = A \mathbf{Z}(t), \qquad \mathbf{Z}(0) = \mathbf{Z}_0. \tag{18}$$

This is easily verified; $e^{A \cdot 0} \mathbf{Z}_0 = I \cdot \mathbf{Z}_0 = \mathbf{Z}_0$ and $(d/dt) eA \cdot t \mathbf{Z}_0 = [A e^{A \cdot t}] \mathbf{Z}_0 = A[e^{A \cdot t} \mathbf{Z}_0]$. For this reason, $e^{A \cdot t}$ is called the fundamental matrix for the equation $\frac{d}{dt} \mathbf{Z}(t) = A \mathbf{Z}(t)$.

From the fact that solutions to (18) are unique, one can also deduce a converse statement:

$$\text{if } \Phi(t) \text{ is a matrix valued solution to} \frac{d}{dt} \Phi(t) = A \Phi(t), \quad \Phi(0) = I, \quad \text{then } \Phi(t) = e^{A \cdot t}. \tag{19}$$

The following basic fact can be proved using either the definition (15) or the fact that $e^{A \cdot t}$ solves equation (16): for any $-\infty < s, t < \infty$, $e^{A \cdot t} e^{A \cdot s} = e^{A \cdot (t+s)}$; in particular, $e^{-A \cdot t} e^{A \cdot t} = e^{A \cdot 0} = I$, and hence e^{-At} is the inverse of e^{At}. However, if $C \neq A$, it is not in general true that $e^{A \cdot t} e^{C \cdot t} = e^{(A+C) \cdot t}$.

Another very useful fact when it come to computing matrix exponentials is the following. Suppose P is an invertible matrix. Observe that

$$[PAP^{-1}]^k = [PAP^{-1}][PAP^{-1}] \cdots [PAP^{-1}] = PA^k P^{-1}.$$

Thus

$$e^{[PAP^{-1}] \cdot t} = I + \sum_{k=1}^{\infty} PA^k P^{-1} \frac{t^k}{k!} = P e^{A \cdot t} P^{-1}.$$

Example 1. Let

$$A = \begin{pmatrix} -\lambda_1 & 0 \\ -\lambda_{21} & -\lambda_2 \end{pmatrix}.$$

Then if $\lambda_1 \neq \lambda_2$,

$$e^{A\cdot t} = \begin{pmatrix} e^{-\lambda_1 t} & 0 \\ \frac{\lambda_{21}}{\lambda_1-\lambda_2}\left(e^{-\lambda_1 t} - e^{-\lambda_2 t}\right) & e^{-\lambda_2 t} \end{pmatrix}. \tag{20}$$

If $\lambda_1 = \lambda_2$,

$$e^{A\cdot t} = \begin{pmatrix} e^{-\lambda_1 t} & 0 \\ \lambda_{21} t e^{-\lambda_1 t} & e^{-\lambda_1 t} \end{pmatrix}. \tag{21}$$

Shreve gives a derivation of these formulas in Lemma 10.2.3 on page 417. It is not necessary to study this derivation in detail. From the characterization of $e^{A\cdot t}$ in (19), it suffice to show that the given formula in each case solves $\frac{d}{dt}\Phi(t) = A\Phi(t)$ with $\Phi(0) = I$. Consider the case of (20). Obviously the given matrix is the identity matrix when $t = 0$. A simple calculation shows that

$$\frac{d}{dt}\begin{pmatrix} e^{-\lambda_1 t} & 0 \\ \frac{\lambda_{21}}{\lambda_1-\lambda_2}\left(e^{-\lambda_1 t} - e^{-\lambda_2 t}\right) & e^{-\lambda_2 t} \end{pmatrix} = \begin{pmatrix} -\lambda_1 e^{-\lambda_1 t} & 0 \\ \frac{-\lambda_{21}}{\lambda_1-\lambda_2}\left(\lambda_1 e^{-\lambda_1 t} - \lambda_2 e^{-\lambda_2 t}\right) & -\lambda_2 e^{-\lambda_2 t} \end{pmatrix}.$$

It is left to the student to show that

$$A\cdot\begin{pmatrix} e^{-\lambda_1 t} & 0 \\ \frac{\lambda_{21}}{\lambda_1-\lambda_2}\left(e^{-\lambda_1 t} - e^{-\lambda_2 t}\right) & e^{-\lambda_2 t} \end{pmatrix} = \begin{pmatrix} -\lambda_1 e^{-\lambda_1 t} & 0 \\ \frac{-\lambda_{21}}{\lambda_1-\lambda_2}\left(\lambda_1 e^{-\lambda_1 t} - \lambda_2 e^{-\lambda_2 t}\right) & -\lambda_2 e^{-\lambda_2 t} \end{pmatrix}.$$

This completes the verification of (20) and (21) may be checked in the same way.

6.3 Multi-dimensional Ito's formula

Let

$$\mu(t) = \begin{pmatrix} \mu_1(t) \\ \vdots \\ \mu_m(t) \end{pmatrix}, \qquad \sigma(t) = \begin{pmatrix} \sigma_{11}(t) & \cdots & \sigma_{1d}(t) \\ \vdots & \vdots & \vdots \\ \sigma_{m1}(t) & \cdots & \sigma_{md}(t). \end{pmatrix},$$

$$W(t) = \begin{pmatrix} W_1(t) \\ \vdots \\ W_d(t) \end{pmatrix}.$$

Let the m-dimensional process X have the dynamics given by

$$dX(t) = \mu(t)dt + \sigma(t)dW_t.$$

Let f be a smooth function that maps $(\mathbb{R}^+, \mathbb{R}^m) \to \mathbb{R}$. Then the process $f(t, X(t))$ has a stochastic differential given by

$$df = \Big\{\frac{\partial f}{\partial t} + \sum_{i=1}^{m} \mu_i(t)\frac{\partial f}{\partial x_i} + \frac{1}{2}\sum_{i,j=1}^{m}(\sigma\sigma^*)_{ij}\frac{\partial^2 f}{\partial x_i \partial x_j}\Big\}dt + \sum_{i=1}^{m}\frac{\partial f}{\partial x_i}\sigma_i(t)\cdot dW_t,$$

where σ_i denotes the ith row f the matrix σ:

$$\sigma_i = [\sigma_{i1}, ..., \sigma_{id}],$$

and σ^* denotes the transpose of σ.

All the expression of df and the partials of f in the above are evaluated at (t, X_t), which we suppressed in the formula for simplicity of notation.

Alternatively, if we denote

$$\begin{aligned} \nabla f &:= (\frac{\partial f}{\partial x_1}, ..., \frac{\partial f}{\partial x_m}) \\ \sigma(t) &= \begin{pmatrix} \frac{\partial^2 f}{\partial x_1^2} & \cdots & \frac{\partial^2 f}{\partial x_1 \partial x_m} \\ \vdots & \vdots & \vdots \\ \frac{\partial^2 f}{\partial x_1 \partial x_m} & \cdots & \frac{\partial^2 f}{\partial x_m^2}(t). \end{pmatrix}, \end{aligned}$$

to be the gradient and the Hessian matrix of f respectively, then the Ito's formula can be written succinctly as

$$df = \Big\{\frac{\partial f}{\partial t} + \mu(t)\cdot\nabla f + \frac{1}{2}tr\big(\sigma\sigma^*(t)H_f\big)\Big\}dt + \nabla f\cdot\sigma dW_t,$$

where $tr(A) := \sum_i A_{ii}$ denotes the trace of a a square matrix A.

6.4 The solution to equation (15)

The matrix exponential function of A may be used to express the solution to (15). This solution is

$$\mathbf{X}(t) = e^{A\cdot t}\mathbf{X}(0) + \int_0^t e^{A\cdot(t-s)}B\,dW(s) + \int_0^t e^{A\cdot(t-s)}c\,ds \tag{22}$$

In the expression $\int_0^t e^{A\cdot(t-s)}B\,dW(t)$, the term $e^{A\cdot(t-s)}B$ is an $m\times d$ matrix multiplying a d-dimensional vector $dW(t)$ of Brownian differentials; hence $\displaystyle\int_0^t e^{A\cdot(t-s)}B\,dW(t)$ is an m-dimensional, vector-valued process.

A variant of (22) is true for representing $\mathbf{X}(T)$ for $T > t$ in terms of T,

$$\mathbf{X}(T) = e^{A\cdot(T-t)}\mathbf{X}(t) + \int_t^T e^{A\cdot(T-s)}B\,dW(s) + \int_t^T e^{A\cdot(t-s)}c(s)\,ds \tag{23}$$

The increments $dW(s)$ for times $s > t$ are independent of $X(t)$; hence (23) exhibits $\mathbf{X}(T)$ is the sum of a linear transformation of $\mathbf{X}(T)$ plus a random vector *independent of* $\mathbf{X}(t)$.

Example 2. Let $\lambda_1 \neq \lambda_2$. Consider

$$d\begin{pmatrix} X_1(t) \\ X_2(t) \end{pmatrix} = \begin{pmatrix} -\lambda_1 & 0 \\ -\lambda_{21} & -\lambda_2 \end{pmatrix}\begin{pmatrix} X_1(t) \\ X_2(t) \end{pmatrix} dt + \begin{pmatrix} \sigma_1 & 0 \\ 0 & \sigma_2 \end{pmatrix}\begin{pmatrix} dW_1(t) \\ dW_2(t) \end{pmatrix} \tag{24}$$

Using the result of Example 1,

$$\begin{aligned}\begin{pmatrix} X_1(t) \\ X_2(t) \end{pmatrix} &= \begin{pmatrix} e^{-\lambda_1 t} & 0 \\ \frac{\lambda_{21}}{\lambda_1-\lambda_2}\left(e^{-\lambda_1 t} - e^{-\lambda_2 t}\right) & e^{-\lambda_2 t} \end{pmatrix}\begin{pmatrix} X_1(0) \\ X_2(0) \end{pmatrix} \\ &\quad + \int_0^t \begin{pmatrix} e^{-\lambda_1(t-s)} & 0 \\ \frac{\lambda_{21}}{\lambda_1-\lambda_2}\left(e^{-\lambda_1(t-s)} - e^{-\lambda_2(t-s)}\right) & e^{-\lambda_2(t-s)} \end{pmatrix}\begin{pmatrix} \sigma_1 dW_1(s) \\ \sigma_2 dW_2(s) \end{pmatrix}.\end{aligned}$$

The student should verify that:

$$\begin{aligned} X_1(t) &= e^{-\lambda t}X_1(0) + \int_0^t e^{-\lambda_1(t-s)}\sigma_1\,dW_1(t) \\ X_2(t) &= \frac{\lambda_{21}}{\lambda_1-\lambda_2}\left(e^{-\lambda_1 t} - e^{-\lambda_2 t}\right)X_1(0) + e^{-\lambda_2 t}X_2(0) \\ &\quad + \int_0^t \frac{\lambda_{21}}{\lambda_1-\lambda_2}\left(e^{-\lambda_1(t-s)} - e^{-\lambda_2(t-s)}\right)\sigma_1\,dW_1(s) \\ &\quad + \int_0^t e^{-\lambda_2(t-s)}\sigma_2\,dW_2(s). \qquad \diamond \end{aligned}$$

To show the validity of (22), write $e^{A\cdot(t-s)} = e^{A\cdot t}e^{-A\cdot s}$ and factor $e^{A\cdot t}$ out of the integrals to write,

$$\mathbf{X}(t) = e^{A\cdot t}\left[\mathbf{X}(0) + \int_0^t e^{-A\cdot s}B\,dW(t) + \int_0^t e^{-A\cdot s}c\,ds\right].$$

Let $Y(t) = \mathbf{X}(0) + \int_0^t e^{-A\cdot s} B\, dW(t) + \int_0^t e^{-A\cdot s} c\, ds$ be the vector-valued Itô process in this expression, and note that $dY(t) = e^{-A\cdot t}\left[c\, dt + B\, dW(t)\right]$. Then

$$\begin{aligned} dX(t) &= \left[\frac{d}{dt} e^{A\cdot t}\right] Y(t)\, dt + e^{A\cdot t} dY(t) \\ &= A e^{A\cdot t} Y(t)\, dt + c\, dt + B\, dW(t) = A\mathbf{X}(t)\, dt + c\, dt + B\, dW(t). \end{aligned}$$

6.5 Joint distribution of the solution to (15)

In Theorem 4.4.9, Shreve states and proves the important fact that the Itô integral of a deterministic integrand is a normal random variable. This fact generalizes. If $W(t)$ is a d-dimensional Brownian motion and $B(t)$ is a deterministic $m \times d$-matrix-valued function, then $\int_0^t B(s)\, dW(s)$ is a normally distributed random vector. Hence, its joint density is determined by its mean vector and covariance matrix. The proof is essentially the same as that of Theorem 4.4.9. This fact has the following consequence for the solution $\mathbf{X}(t)$ of (15): *for any $0 \le t \le T$, the conditional distribution of $X(T)$ given $X(t)$ is Gaussian (normal).* In particular, if $\mathbf{X}(0)$ is deterministic or is a normal random variable independent of W, then $\{\mathbf{X}(t);\, t \ge 0\}$ is a vector-valued, Gaussian process.

Exercise 10.1 in Shreve is about the mean vector and covariance matrix of the process defined in Example 2.

6.6 How (15) changes under affine change of variable

Let $\mathbf{X}$ solve equation (15), let P be an invertible $m \times m$ matrix, let a be an m-vector, and define

$$\mathbf{Y}(t) = P\mathbf{X}(t) + a.$$

Then $\mathbf{Y}(t)$ also satisfies a system of linear stochastic differential equations. Note that $\mathbf{X}(t) = P^{-1}(\mathbf{Y}(t) - a)$. Thus,

$$\begin{aligned} d\mathbf{Y}(t) = P\, d\mathbf{X}(t) &= P\left[(A\mathbf{X}(t) + c)\, dt + B\, dW(t)\right] \\ &= \left[PAP^{-1}\mathbf{Y}(t) + P(c - AP^{-1}a)\right] dt + PB\, dW(t). \end{aligned}$$

Linear transformations like this are extremely useful for simplifying linear systems. For example, if A is invertible, and we choose $a = PA^{-1}C$, then $(c - AP^{-1}a)$ equals the zero vector, and hence $d\mathbf{Y}(t) = PAP^{-1}\mathbf{Y}(t)\,dt + PB\,dW(t)$. More importantly, one can choose P so that the matrix PAP^{-1} has a canonical form that is simple to work with, in terms of calculating $e^{PAP^{-1}\cdot t} = Pe^{A\cdot t}P^{-1}$ and in terms of understanding how the different components of $Y_i(t)$ influence one another. For example, if A has a basis of eigenvectors with real eigenvalues $\lambda_1, \ldots, \lambda_m$, P can be chosen so that PAP^{-1} is the diagonal matrix

$$PAP^{-1} = \begin{pmatrix} \lambda_1 & 0 & 0 & \cdots & 0 \\ 0 & \lambda_2 & 0 & \cdots & 0 \\ \vdots & & \ddots & & \vdots \\ 0 & \multicolumn{3}{c}{\dots\dots\dots\dots} & \lambda_m \end{pmatrix}.$$

It is easily seen that

$$e^{PAP^{-1}\cdot t} = \begin{pmatrix} e^{\lambda_1 t} & 0 & 0 & \cdots & 0 \\ 0 & e^{\lambda_2 t} & 0 & \cdots & 0 \\ \vdots & & \ddots & & \vdots \\ 0 & \multicolumn{3}{c}{\dots\dots\dots\dots} & e^{\lambda_m t} \end{pmatrix}.$$