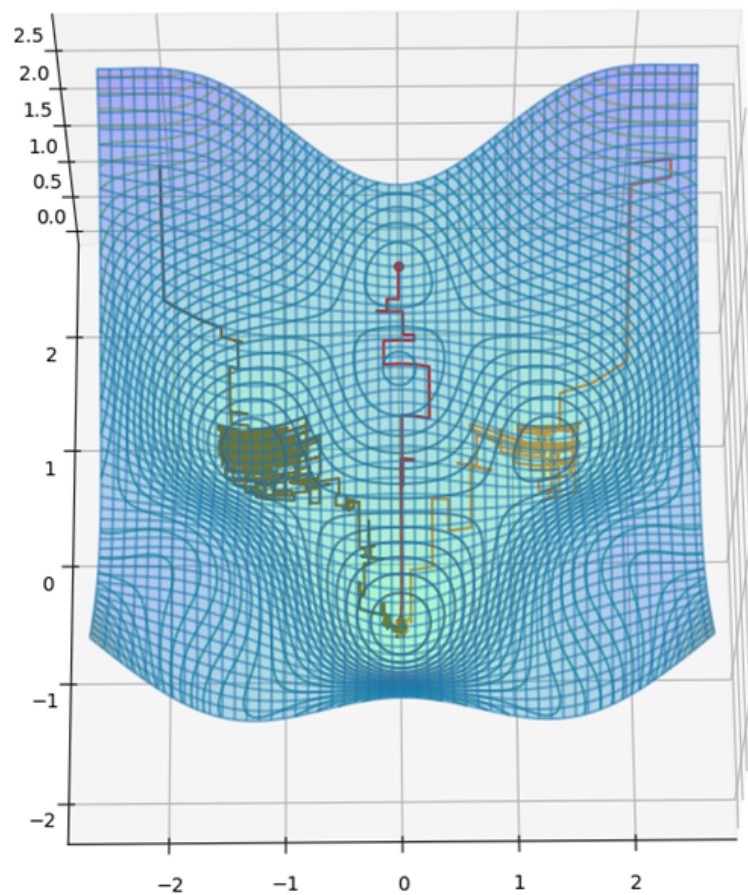


MASTEROPPGAVE:

**Konstruksjon av alternativ stokastisk gradientalgoritmen  
ved bruk av stokastisk analyse og optimal kontrollteori.**

Bjarte Beil-Myhre



VEILEDER:

Torstein Kastberg Nilssen

**Universitetet i Agder, 2023**

Fakultet for teknologi og realfag  
Institutt for matematiske fag

## Forord

Denne oppgaven avslutter mine 4 år på UIA. Det har vært en spennende reise med mange morsomme utfordringer. Jeg har lært ting jeg ikke ante eksisterte og fått et glimt inn i den enorme verden av matematikk. Jeg sitter igjen med stor respekt for alle de kloke hodene som har produsert denne fascinerende kunnskapen. Og ikke minst takknemlighet for at jeg har fått muligheten til å ta del i noe av den.

Jeg vil rette en stor takk til min kjære familie for støtten gjennom disse 4 årene og til Torstein Kastberg Nilssen som har veiledet meg og hjulpet meg underveis i arbeidet med denne oppgaven.

Borøya, mai 2023

## Abstrakt

Oppgaven viser at den stokastiske gradientalgoritmen med en minibatch størrelse  $r$ :

$$x_{k+1} = x_k - \eta \frac{1}{r} \sum_{i=1}^n \nabla f_i(x_k) 1_{\delta(k+1)}(i)$$

konvergerer i fordeling, når læringsraten  $\eta \rightarrow 0$  mot løsningen av den stokastiske differensialligningen:

$$\begin{cases} dX_t = -\nabla f(X_t)dt + (\bar{\eta}\Sigma(X_t))^{\frac{1}{2}}dB_t & t > 0 \\ X_0 = x_0 & x_0 \in \mathbb{R}^d \end{cases}$$

Basert på dette resultatet skal vi konstruere av en alternativ stokastisk gradientalgoritme ved bruk av stokastisk analyse og kontrollteori.

# Abstract

This thesis shows that the stochastic gradient algorithm, with a minibatch size  $r$ :

$$x_{k+1} = x_k - \eta \frac{1}{r} \sum_{i=1}^n \nabla f_i(x_k) 1_{\delta(k+1)}(i)$$

converges in distribution, as the learning rate  $\eta \rightarrow 0$  to a solution of the stochastic differential equation:

$$\begin{cases} dX_t = -\nabla f(X_t)dt + (\bar{\eta}\Sigma(X_t))^{\frac{1}{2}}dB_t & t > 0 \\ X_0 = x_0 & x_0 \in \mathbb{R}^d \end{cases}$$

Based on this result, we aim to construct an alternative stochastic gradient algorithm using stochastic analysis and control theory.

# Innhold

<b>Forord</b>	<b>2</b>
<b>Abstrakt</b>	<b>3</b>
<b>Abstract</b>	<b>4</b>
<b>Notasjon</b>	<b>7</b>
<b>Innledning</b>	<b>8</b>
<b>1 Innledende teori</b>	<b>10</b>
1.0.1 Motivasjon . . . . .	10
1.1 Sannsynlighetsteori . . . . .	11
1.1.1 Betinget forventning . . . . .	13
1.2 Brownsk bevegelse . . . . .	14
1.2.1 Integrasjon med hensyn på den Brownske bevegelsen $B_t$ . . . . .	15
1.2.2 Itô's formel . . . . .	17
1.2.3 Martingaler . . . . .	19
1.2.4 Diffusjonsprosesser . . . . .	20
1.2.5 Ornstein-Uhlenbeck prosess . . . . .	21
1.3 Sammenheng mellom PDE-er og SDE-er . . . . .	22
1.3.1 Generatoren til diffusjoner . . . . .	22
1.3.2 Fokker-Planck-ligningen tilknyttet Itô-diffusjonsprosesser	24
<b>2 Optimal kontrollteori</b>	<b>30</b>
2.1 Pontryagins maksimalprinsipp . . . . .	32
2.1.1 Optimal investeringsstrategi for $t \in [0, T]$ . . . . .	32
2.1.2 Pontryagins maksimalprinsipp for fiksert $T$ . . . . .	32
2.1.3 Løsning for den optimale investeringsstrategi . . . . .	33
2.1.4 Dynamisk programmering . . . . .	34
2.2 Hamilton-Jacobi-ligningen . . . . .	36
2.2.1 Konstruksjon av den optimale kontrollen $\alpha^*(\cdot)$ . . . . .	38
2.2.2 Bevis for at $\alpha^*(\cdot)$ er optimal . . . . .	38

<b>3</b>	<b>Gradient nedstigning</b>	<b>39</b>
3.1	Konvergens av stokastisk gradientalgoritmen . . . . .	40
3.1.1	Egenskaper ved den stokastiske gradient nedstigningen	41
3.1.2	Et generelt funksjonalgrenseteorem . . . . .	44
3.2	Bevis teorem 3.2 (for en generell stokastisk gradientalgoritme)	45
3.3	Alternativ prosess $\bar{X}_t$ . . . . .	49
<b>4</b>	<b>Konstruksjon av algoritme</b>	<b>51</b>
4.0.1	Innføring av kontroll i den stokastiske gradientskjema	52
4.1	Løsning kontroll problem . . . . .	53
4.2	Overgang til $\mathbb{R}^d$ med eksempel . . . . .	56
4.2.1	Implementering . . . . .	57
4.2.2	Resultater av test . . . . .	58
<b>5</b>	<b>Oppsummering</b>	<b>62</b>
<b>6</b>	<b>Referanser</b>	<b>63</b>

# Notasjon

$\mathbb{R}^d$  Euklidsk rom i  $d$  dimensjoner

$C_0^2(\cdot)$  Den første og andrederiverte er kontinuerlig og har kompakt støtte

$F[f(\cdot)]$  med klammer funksjonal

$\mathbb{E}[\cdot]$  = Forventning

$\mathcal{F}[\cdot](\xi)$  = Fourier transformen på  $\mathbb{R}^n$

n.o. nesten overalt

n.s. nesten sikkert

$S_d$  rommet med alle  $d \times d$  symmetriske matriser

$\mathcal{B}(\mathbb{R})$  er Borel  $\sigma$ -algebra på  $\mathbb{R}$

$\|\cdot\| : \mathbb{R}^n \rightarrow [0, \infty)$  Euklidske normen.  $x \in \mathbb{R}^n$   $\|x\| = \sqrt{x_1^2 + x_2^2 \dots x_n^2}$

$Var(X)$  er variansen til  $X$

$X \sim \mathcal{N}(\mu, \sigma)$   $X$  er normalfordelt med forventning  $\mu$  og varians  $\sigma$

$\int |f(x)|dx < \infty$   $f$  er integrerbar

# Innledning

Målet for denne oppgaven er å vise at når læringsraten  $\eta \rightarrow 0$  konvergerer den stokastiske gradientalgoritmen,

$$x_{k+1} = x_k - \eta \frac{1}{r} \sum_{i=1}^n \nabla f_i(x_k) 1_{\delta(k+1)}(i)$$

mot løsningen av den en stokastisk differensialligning

$$\begin{cases} dX_t = -\nabla f(X_t)dt + (\bar{\eta}\Sigma(X_t))^{\frac{1}{2}}dB_t & t > 0 \\ X_0 = x_0 & x_0 \in \mathbb{R}^d \end{cases}$$

Dette resultatet skal vi bruke, ved hjelp av stokastisk analyse og optimal kontrollteori, til å konstruere en alternativ stokastisk gradientalgoritmen hvor vi kan kontrollere læringsraten når vi nærmer oss minimum.

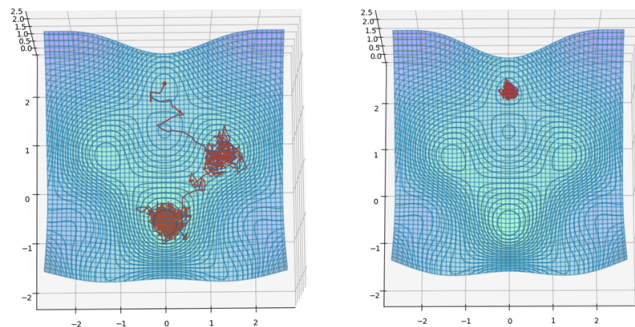
Utgangspunktet kommer fra minimeringsproblemer vi kan skrive på formen:

$$\min_{x \in \mathbb{R}^d} f(x) = \min_{x \in \mathbb{R}^d} \frac{1}{n} \sum_{i=1}^n f_i(x)$$

Dette er gjerne problemer vi finner igjen innenfor maskinlæring hvor vi ønsker å minimere en tilhørende kostfunksjon, hvor  $n$  er antallet data/målepunkter og  $d$  er antallet parametere vi minimerer over. Kompleksiteten og antallet dimensjoner i disse problemene gjør at det ikke hensiktsmessig å bruke den vanlige gradient algoritmen, og i stedet brukes en form for stokastiske gradientalgoritmen.

Her følger man ikke hele gradienten, altså gradienten til alle datapunktene  $n$  på en gang, men for hvert steg trekker vi  $r$  tilfeldig valgte datapunkter og følger gradienten til disse.

Læringsraten  $\eta$  er en valgt lengden på stegene i hver iterasjon, og i praksis påvirker dette valget av  $\eta$  i stor grad hvor godt algoritmen presterer. Vi skal kun forholde oss til standard gradient algoritmen.



Figur 1: Stokastisk gradientalgoritme med ulike læringsrater på funksjonen  $f(x, y) = \sin(2y) \frac{\cos(2x)}{2} + \frac{x^2+y^2}{5} + 0.4$



Plottene i figur 1 viser eksempler på bruk av to ulike læringsrater i den stokastiske gradientalgoritmen. I tilfellet til høyre er læringsraten lav og vi kommer oss ikke vekk fra startpunktet, mens til venstre kommer vi oss vekk, men algoritmen er preget av mye støy.

Vi skal i denne oppgaven vise hvordan vi kan lage en tilpasset algoritmen som på sett og vis er en kombinasjon av begge. Først ønsker vi å ta store steg som til venstre og så når vi nærmer oss minimum kan vi kontrollere  $\eta$  slik at vi stopper gradvis opp rundt minimum med mindre og mindre steg.

Opgave er delt inn i fire deler.

**Del 1:** Den første delen inneholder sannsynlighetsteori og hovedresultater fra stokastisk analyse. Samt en gjennomgang av sammenhengen mellom fordelingen til diffusjonsprosesser og partielle differensialligninger gjennom Fokker-Planck-ligningen.

**Del 2:** Her gir vi en kort innføring i optimal kontroll teori, med eksempler på bruk av Pontryagins maksimalprinsipp og bevis av Hamilton-Jacobi-ligningen. Hovedfokuset har vært å gi en intuisjon av hva kontrollteori.

**Del 3:** Denne delen inneholder hovedresultatene i oppgaven. Her går vi gjennom bevis for at den stokastiske gradientalgoritmen, med minibatch størrelse  $r$  konvergerer i fordeling mot løsningen av en stokastisk differensialligning. Dette beviset er en generalisering av Lanconelli og Laurias bevis (Lanconelli, & Lauria, 2022).

Vi ser og på en alternativ stokastisk differensialligning hvor vi viser, ved bruk av Fokker-Planck-ligningen at løsningen har samme fordeling.

**Del 4:** I den fjerde og siste delen er det en gjennomgang av argumentasjonen bak konstruksjonen av den tilpassende stokastisk gradientalgoritmen til Li, Tai og E (Li et al., 2017). I tillegg er algoritmen implementert på et konstruert eksempel hvor den tilpassede stokastiske gradientalgoritmen er testet mot den stokastiske gradientalgoritmen.

# 1 Innledende teori

Vi skal nå gjennomgå en del resultater hentet fra stokastisk analyse. Fremstillingen følger i hovedsak (Øksendal, 2013, kapittel 2-5 og 7), men det er og deler fra Varadhans bok på sannsynlighetsteori (Varadhan, 2012, kapittel 1 og 4) og (Evans, 2013, kapittel 1).

## 1.0.1 Motivasjon

Det første vi ønsker å gjøre er å gi mening til hva en stokastisk differensialligning (SDE) er og hvordan vi skal tolke de ulike leddene i ligningen. Etter vi har gjort dette skal vi se på Itô's formel. Før går vi nærmere inn på klassen tids-homeogene Itô diffusjonsprosesser og ser på sammenhengen mellom løsningen av disse prosessene og en tilhørende PDE, kalt Fokker-Planck-ligningen.

Vi starter med en ordinær differensialligning for et fiksert punkt  $x_0 \in \mathbb{R}^n$ :

$$\begin{cases} \dot{x}(t) = f(x(t)) & t > 0 \\ x(0) = x_0 \end{cases}$$

$f : \mathbb{R}^n \rightarrow \mathbb{R}^n$  er en glatt funksjon og for hver  $x(0) = x_0$  har vi løsningen  $x(\cdot) : [0, \infty) \rightarrow \mathbb{R}^n$  som vi kaller banen gitt  $x_0$ . Denne beskrivelsen av et fenomen kan fungere teoretisk, men når vi i praksis starter å måle i den virkelige verden er ikke alt like glatt og forutsigbart. Det kan f.eks. være støy som påvirker prosessen vi observerer så for å modellere usikkerheten i systemet må vi legge til støy  $\xi(t)$ :

$$\begin{cases} \dot{X}(t) = f(X(t), t) + \sigma(X(t), t)\xi(t) & t > 0 \\ X(t_0) = x_0 \end{cases} \quad (1)$$

$\sigma : \mathbb{R}^{n \times m} \rightarrow \mathbb{R}^n$  og  $\xi(t)$  er m-dimensjonal hvit støy. Velger vi nå  $f(X_t, t) = 0$  og  $\sigma = 1$  for alle t, får vi Wiener prosessen eller den Brownske bevegelsen notert  $B(\cdot)$ :  $\dot{B}(t) = \xi(t)$  Vi setter dette tilbake i (1):

$$\frac{d}{dt}X(t) = f(X(t), t) + \sigma(X(t), t)\frac{dB(t)}{dt}$$

Nå er ikke den Brownske bevegelsen deriverbar i noe punkt så uttrykket over gir egentlig ikke mening, men når vi multipliserer på begge sider med dt har vi kommet frem til den stokastiske differensialligningen.

$$SDE : \begin{cases} dX(t) = f(X(t), t)dt + \sigma(X(t), t)dB(t) & t > 0 \\ X(0) = x_0 \end{cases} \quad (2)$$

Den tilhørende løsningen får vi ved å integrere m.h.p. t:

$$X(T) = X(0) + \int_0^T b(X(t), t)dt + \int_0^T \sigma(X(t), t)dB(t) \quad (3)$$

For å finne en definisjon på en SDE og en tilhørende løsning trenger vi svar på hvordan vi skal tolke  $\int_0^T \sigma(X(t), t) dB(t)$ . Fra vanlig kalkulus vet vi at når vi integrerer over en funksjon  $f(t)$  kan  $df(t)$  erstattes med  $f'(t)dt$  slik at,  $\int_0^t g(s)df(s) = \int_0^t g(s) \cdot f'(s)ds$ . Denne overgangen krever dog at  $f$  er deriverbar.  $B(t)$  er ikke deriverbar i noe punkt så løsningen av  $\int_0^T \sigma(X(t), t)dB(t)$  gir ikke mening med tilnærmingen vi har i ordinær kalkulus. Vi skal nå gå gjennom oppbygningen av dette integralet slik at dette gir mening.

## 1.1 Sannsynlighetsteori

Den stokastiske differensialligningen består av noe tilfeldig og er dermed knyttet til et utfallsrom  $\Omega$ , med utfallene  $\omega \in \Omega$ . Vi kan generelt definere et sannsynlighetsrom med utgangspunkt i utfallsrommet  $\Omega$ .

**Definisjon 1.1** (Sannsynlighetsrom). *Et sannsynlighetsrom består av trippe-len  $(\Omega, \mathcal{F}, P)$  hvor:*

(i)  $\Omega$  er utfallsrommet

(ii)  $\mathcal{F}$  er en  $\sigma$ -algebra på  $\Omega$

(iii)  $P$  er en avbildning fra  $\mathcal{F}$  inn i  $[0, 1]$  hvor  $P(\emptyset) = 0$  og  $P(\Omega) = 1$  og  $P(\cup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} P(A_i)$  når  $A_i \cap A_j = \emptyset, \forall j \neq i. A_i \in \mathcal{F}$

La  $\mathcal{B}(\mathbb{R}^n)$  være Borel  $\sigma$ -algebraen på  $\mathbb{R}^n$ . Vi sier at en funksjon  $X : \Omega \rightarrow \mathbb{R}^n$  er målbar dersom for alle  $B \in \mathcal{B}(\mathbb{R}^n)$  så er  $X^{-1}(B) \in \mathcal{F}$ , hvor  $\mathcal{F}$  er en  $\sigma$ -algebra på  $\Omega$ .

En tilfeldig variabel  $X : \Omega \rightarrow \mathbb{R}^n$  er en  $\mathcal{F}$ -målbar funksjon, og hver tilfeldig variabel  $X$  på  $(\Omega, \mathcal{F})$  genererer et sannsynlighetsmål  $\mu_X$  på  $\mathbb{R}^n$ . Dette målet  $\mu_X$  er knyttet til avbildningen  $P$  på  $\mathcal{F}$  ved at for alle  $B \in \mathcal{B}(\mathbb{R}^n)$  så er  $\mu_X(B) = P(X^{-1}(B))$ .

**Definisjon 1.2** (Indikatorfunksjonen). *La  $A \in \mathcal{F}$  da er indikatorfunksjonen til  $A$  en tilfeldig variabel,*

$$1_A(\omega) = \begin{cases} 1 & : \omega \in A \\ 0 & : \omega \notin A \end{cases}$$

Vi kaller en tilfeldig variable  $X : \Omega \rightarrow \mathbb{R}$  enkel dersom:

$$X(\omega) = \sum_{i=1}^n a_i 1_{A_i}(\omega)$$

hvor  $A_i \in \mathcal{F}$  for alle  $i$ ,  $a_i \in \mathbb{R}$  og  $\cup A_i = \Omega$ . Vi kan nå definere forventingen til en tilfeldig variabel  $X$  :

**Definisjon 1.3** (Forventning).

$$\mathbb{E}[X(\cdot)] = \int_{\Omega} X(\omega) dP(\omega) \quad (4)$$

$\Omega$  er utfallsrommet og  $P$  er den tilhørende sannsynlighetsmålet til  $X$  på  $\Omega$ .

for den enkle funksjonen  $X(\omega)$  får vi forventningen:

$$\mathbb{E}[X(\cdot)] = \int_{\Omega} \left( \sum_{i=1}^n a_i 1_{A_i}(\omega) \right) dP(\omega) = \sum_{i=1}^n \int_{\Omega} a_i 1_{A_i}(\omega) dP(\omega) = \sum_{i=1}^n a_i P(A_i)$$

Fra definisjonen av forventning integrerer vi over utfallsrommet  $\Omega$  og ikke over  $\mathbb{R}^n$ . Vi kan i noen tilfeller bytte fra målet  $P$  på  $\Omega$  over til Lebesgue-målet på  $\mathbb{R}^n$ .

For at vi skal kunne gjøre denne overgangen krever det at  $P$  er absolutt kontinuerlig med hensyn på Lebesgue-målet.

**Definisjon 1.4** (Absolutt kontinuerlig). *La både  $\nu$  og  $\lambda$  være mål på  $\mathcal{F}$ . Vi sier målet  $\nu$  er absolutt kontinuerlig med hensyn på  $\lambda$ , notert:  $\nu \ll \lambda$ , dersom:*

$$\forall A \in \mathcal{F}, \quad \text{hvor} \quad \nu(A) = 0 \implies \lambda(A) = 0$$

En konsekvens av dette er Radon-Nikodym teoremet:

**Teorem 1.1** (Radon-Nikodym). *Dersom målet  $\nu \ll \mu$ , da finnes det en integrerbar funksjon  $f(\omega)$  slik at:*

$$\nu(A) = \int_A f(\omega) d\mu$$

for alle  $A \in \mathcal{F}$ . Denne funksjonen  $f$  er unik, nesten overalt, og kalles den Radon-Nikodym deriverte til  $\nu$  med hensyn på  $\mu$ . Og vi kan skrive dette:

$$f(\omega) = \frac{d\nu}{d\mu}$$

Fra Radon-Nikodym teoremet får vi at dersom sannsynlighetsmålet  $P$  på  $\Omega$  er absolutt kontinuerlig m.h.p. Lebesgue målet  $\mu$  på  $\mathbb{R}^n$  finnes det en nesten overalt unik  $f$  slik at:

$$\frac{dP}{d\mu} = f(x) \implies dP = f(x) d\mu$$

$f(x)$  er sannsynlighetstettheten knyttet til den tilfeldige variabelen som gir oss overgangen mellom de to målene. Når den Radon-Nikodym deriverte finnes kan vi skrive om forventningen:

$$\mathbb{E}(X(\cdot)) = \int_{\Omega} X(\omega) dP(\omega) = \int_{\mathbb{R}^n} x f(x) d\mu(x) \quad x \in \mathbb{R}^n \quad (5)$$

### 1.1.1 Betinget forventning

Betinget forventning er knyttet til hvilken informasjon vi allerede har eller hvilke betingelser vi ønsker å begrense oss til.

**Eksempel 1.1.** *Du har to terninger, en rød og en gul, og triller en av dem. Forventningen er 3.5 for et terningkast, men dersom du vet at du har fått partall er forventningen 4. Fordelingen endrer seg på bakgrunn av informasjonen siden sannsynligheten for å få 1,3,5 nå er 0. Utfallet er altså avhengig av informasjonen vi har. Dersom vi får informasjonen om at terningen er rød påvirker ikke dette forventningen til utfallet. Utfallet er dermed uavhengig av informasjonen om fargen.*

Vi skal nå definere den betingete forventningen. La  $(\Omega, \mathcal{F}, P)$  være et sannsynlighetsrom og  $X$  er en tilfeldig variabel fra  $\Omega \rightarrow \mathbb{R}^n$  slik at  $\mathbb{E}[|X|] < \infty$ . Dersom  $\mathcal{H}$  er en  $\sigma$ -algebra hvor  $\mathcal{H} \subset \mathcal{F}$ . Da er den betingede forventningen til  $X$  gitt  $\mathcal{H}$ , notert  $\mathbb{E}[X|\mathcal{H}]$  gitt ved:

**Definisjon 1.5** (Betinget forventning).  $\mathbb{E}[X|\mathcal{H}]$  er en n.s. entydig funksjon fra  $\Omega \rightarrow \mathbb{R}^n$  som tilfredstiller:

$$\begin{aligned} 1 : & \mathbb{E}[X|\mathcal{H}], \quad \text{er } \mathcal{H}\text{-målbar} \\ 2 : & \int_H \mathbb{E}[X|\mathcal{H}] dP = \int_H X dP, \quad \text{for alle } H \in \mathcal{H} \end{aligned} \tag{6}$$

Vi sier forventningen til  $X$  gitt  $\mathcal{H}$ . Fra definisjon av betinget forventning følger 5 egenskaper vi trenger videre:

- 1 :  $\mathbb{E}[aX + bY|\mathcal{H}] = a\mathbb{E}[X|\mathcal{H}] + b\mathbb{E}[Y|\mathcal{H}]$  (Linearitet)
- 2 :  $\mathbb{E}[\mathbb{E}[X|\mathcal{H}]] = \mathbb{E}[X]$  (Tårnegenskapen)
- 3 :  $\mathbb{E}[X|\mathcal{H}] = X$ , Hvis  $X$  er  $\mathcal{H}$ -målbar.
- 4 :  $\mathbb{E}[X|\mathcal{H}] = \mathbb{E}[X]$ , Hvis  $X$  er uavhengig av  $\mathcal{H}$
- 5 :  $\mathbb{E}[Y \cdot X|\mathcal{H}] = Y \cdot \mathbb{E}[X|\mathcal{H}]$  Hvis  $Y$  er  $\mathcal{H}$ -målbar;  $\cdot$  er indreproduktet på  $\mathbb{R}^n$

Vi skal nå definere en stokastisk prosess. La  $T \in (0, \infty)$ .

**Definisjon 1.6** (Stokastisk prosess). *En stokastisk prosess  $X$  er en parametrisert samling tilfeldige variabler  $X = \{X_t\}_{t \in [0, T]}$  hvor alle  $X_t$  er definert på et felles sannsynlighetsrom  $(\Omega, \mathcal{F}, P)$  og tar verdier i  $\mathbb{R}^n$ .*

Vi kan betrakte en stokastisk prosess  $X(\omega, t)$  på to ulike måter. Dersom vi fikserer  $t$ ,  $X(\cdot, t)$  har vi en tilfeldig variabel. Fikserer vi  $\omega$ ,  $X(\omega, \cdot)$  har vi et eksempel på en sti.

**Eksempel 1.2.** Hver dag i en uke skal vi kaste en mynt og noterer ned hva vi får. Kron gir 1 og mynt gir 0. I slutten av uken skal vi summere opp hva vi har fått tilsammen. Dette er en stokastisk prosess  $\{X_t\}_{t \in [1, \dots, 7]}$  hvor vi for hver dag  $t$  har vi en tilfeldig variabel  $X_t : \Omega \rightarrow \{0, 1\}$ , enten kron eller mynt.

I eksempel 1.2 er informasjon om hva som har skjedd de foregående dagene ikke relevant for utfallet i dag, det påvirker ikke sannsynligheten for å få kron eller mynt i dag. Denne uavhengigheten gjelder langt fra alle prosesser så vi trenger et rammeverk for å forholde oss til informasjon vi allerede har knyttet til stokastiske prosesser.

**Definisjon 1.7** (Filtrasjon). La  $(\Omega, \mathcal{F}, P)$  være et sannsynlighetsrom og  $\mathcal{F}_t \subseteq \mathcal{F}$  for alle  $t \in [0, T]$  være del- $\sigma$ -algebraer. Da er  $\mathcal{F}_T = \{\mathcal{F}_t\}_{t \in [0, T]}$  en filtrasjon dersom  $\mathcal{F}_s \subseteq \mathcal{F}_t \quad \forall s, t \quad \text{s.a.} \quad s \leq t$

Filtrasjonen  $\mathcal{F}_t$  tenker vi på som informasjonen vi har opp til tiden  $t$ . Der som en tilfeldig prosess er uavhengig av filtrasjonen  $\mathcal{F}_s$  vil det bety at utfallet ikke påvirkes av den ekstra informasjonen og

$$E[X(\cdot, t) | \mathcal{F}_s] = E[X(\cdot, t)] \quad t > s$$

Vi kan ha en filtrasjon  $\mathcal{F}_t$  som inneholder all informasjon om prosessen frem til tiden  $t$ , og har et eget begrep for tilfellene hvor dette holder for alle  $t$ .

**Definisjon 1.8** (Adaptert). En stokastiske prosess  $X_t$  er adaptert til filtrasjonen  $\{\mathcal{F}_t\}_{t \in T}$  hvis for alle  $t \in T$  så er  $X_t$  målbar m.h.p.  $\mathcal{F}_t$ .

## 1.2 Brownsk bevegelse

En Brownsk bevegelse er en stokastisk prosess oppkalt etter botanikeren Robert Brown som skal ha observert og beskrevet hvordan et pollen i vann blir dyttet rundt av andre mindre partikler i vannet. Denne banen virket tilfeldig og var ikke mulig å forutse. I dag når vi snakker om Brownsk bevegelse er det snakk om en stokastisk prosess med gitte egenskaper.

**Definisjon 1.9** (Browsk bevegelse). En Brownsk bevegelse  $B_t$  er en stokastisk prosess  $(0, \infty) \times \Omega \rightarrow \mathbb{R}^n$  hvor  $B_t$  tilfredstiller:

- (i)  $B_0 = 0$
- (ii) Den brownske bevegelsen er gaussisk  $B_t \sim \mathcal{N}(0, I_{n \times n} t)$   
 $I_{n \times n}$  er  $n \times n$  identitetsmatrise.
- (iii) Uavhengige inkrementer:  $\forall t \geq s \quad B_t - B_s$  er uavhengig av filtrasjonen  $\mathcal{F}_s$

**Merknad 1.** Den brownske bevegelsene er ikke nødvendigvis kontinuerlig, men fra Kolmogorov-Centsov teoremet<sup>1</sup> kan man vise at det alltid eksisterer en modifikasjon  $\tilde{B}_t$  av den Brownske bevegelsen  $B_t$ , som er  $P$ -nesten sikkert kontinuerlig. Videre vil det gjennomgående refereres til den kontinuerlige modifikasjonen om ikke annet er spesifisert.

---

<sup>1</sup>Øksensdal

Fra definisjon av den Brownske bevegelsen kan vi trekke ut et par sentrale resultater. La  $\{B_t\}_{t \in [0, T]}$  være adaptert til filtrasjonen  $\{\mathcal{F}_t\}_{t \geq 0}$ . Da får vi fra uavhengighet (iii) til den Brownske bevegelsen og den betinget forventning gitt  $\mathcal{F}_s$  og  $s < t \leq T$  :

$$E[B_t - B_s | \mathcal{F}_s] = E[B_t - B_s] = E[B_t] - E[B_s] = 0$$

Og siden forventning er en lineær operator og  $B_s$  er målbar med hensyn på  $\mathcal{F}_s$ :

$$E[B_t | \mathcal{F}_s] = E[(B_t - B_s) + B_s | \mathcal{F}_s] = E[(B_t - B_s) | \mathcal{F}_s] + E[B_s | \mathcal{F}_s] = E[B_t - B_s] + B_s = B_s$$

### 1.2.1 Integrasjon med hensyn på den Brownske bevegelsen $B_t$

Vi skal ikke gå inn i alle detaljene, men gi skisse av oppbygningen av Itô-integralet  $\int_0^T \phi(s, \omega) dB_s(\omega)$  slik at den stokastiske differensialligningen (3) gir mening.

La sannsynlighetsrommet  $(\Omega, \mathcal{F}, P)$  være gitt med en tilhørende filtrasjon  $\{\mathcal{F}_t\}_{t \in [0, T]}$ .  $B_t$  er en Brownsk bevegelse målbar med hensyn på filtrasjonen  $\{\mathcal{F}_t\}_{t \in [0, T]}$  på  $(\Omega, \mathcal{F}, P)$ .

**Definisjon 1.10.** la  $\mathcal{V} = \mathcal{V}(S, T)$  være klassen med funksjoner

$$\phi : [0, \infty) \times \Omega \rightarrow \mathbb{R}$$

slik at

**1:**  $(t, \omega) \rightarrow \phi(t, \omega)$  er  $\mathcal{B} \times \mathcal{F}$  målbar.  $\mathcal{B}$  er Borel- $\sigma$ -algebaren på  $[0, \infty)$

**2:**  $\phi(t, \omega)$  er adaptert til filtrasjonen  $\mathcal{F}_t$

**3:**  $\mathbb{E}[\int_S^T \phi(t, \omega)^2 dt] < \infty$

En tilfeldig prosess  $\xi$  er elementær dersom  $\xi(t, \omega) = \sum_{i=1} \xi_i(t, \omega) 1_{[t_i, t_{i+1})}(t)$  hvor  $\xi_i : \Omega \rightarrow \mathbb{R}$  er en tilfeldig variabel som er adaptert til filtrasjonen  $\mathcal{F}_{t_i}$ . Videre antar vi og  $\xi \in \mathcal{V}$ . Vi definerer nå integralet med hensyn på den Brownske bevegelsen:

$$\int_0^T \xi(s, \omega) dB_s(\omega) = \sum_{i=1} \xi_i(B_{t_{i+1}} - B_{t_i})$$

Vi kan nå tilnærme  $\phi(t, \omega) \in \mathcal{V}$  med funksjonen  $\xi(t, \omega) = \sum_{i=1} \xi_i 1_{[t_i, t_{i+1})}(t)$  hvor vi velger  $\xi_i = \phi(t_i)$  på intervallet  $[t_i, t_{i+1})$ . Og da tilnærme Itô-integralet ved:

$$\int_0^T \phi(s, \omega) dB_s(\omega) = \sum_{i=1} \xi_i(B_{t_{i+1}} - B_{t_i}) \quad (7)$$

Som en direkte konsekvens av Itô-integralet og egenskapene til den Brownske bevegelsen kan vi utlede Itô-Isometrien:

**Lemma 1.2** (Itô isometri).  $\phi(t, \omega)$  er bregrenset og elementær funksjon. Da er Itô isometrien:

$$E \left[ \left( \int_0^T \phi(s, \omega) dB_s(\omega) \right)^2 \right] = E \left[ \int_0^T \phi(s, \omega)^2 ds \right] \quad (8)$$

*Bevis.*

$$\begin{aligned} E \left[ \left( \int_0^T \phi(s, \omega) dB_s(\omega) \right)^2 \right] &= \mathbb{E} \left[ \left( \sum_{i=1}^m \xi_i (B_{t_{i+1}} - B_{t_i}) \right) \left( \sum_{i=1}^m \xi_i (B_{t_{i+1}} - B_{t_i}) \right) \right] \\ &= \mathbb{E} \left[ \sum_{i=j=1}^m \xi_i^2 (B_{t_{i+1}} - B_{t_i})^2 + 2 \sum_{i>j=1}^m \xi_i \xi_j (B_{t_{i+1}} - B_{t_i}) (B_{t_{j+1}} - B_{t_j}) \right] \\ &= \mathbb{E} \left[ \sum_{i=1}^m \xi_i^2 (B_{t_{i+1}} - B_{t_i})^2 \right] + 2 \mathbb{E} \left[ \sum_{i>j=1}^m \xi_i \xi_j (B_{t_{i+1}} - B_{t_i}) (B_{t_{j+1}} - B_{t_j}) \right] \\ &= \sum_{i=1}^m \mathbb{E} \left[ \xi_i^2 (B_{t_{i+1}} - B_{t_i})^2 \right] + 2 \sum_{i>j=1}^m \mathbb{E} \left[ \xi_i \xi_j (B_{t_{i+1}} - B_{t_i}) (B_{t_{j+1}} - B_{t_j}) \right] \end{aligned}$$

Ettersom  $\xi_i$  og  $\xi_j$  er målbar m.h.p filtrasjonen  $\mathcal{F}_{t_i}$  og bruk av egenskap 2 til den betingede forventningen.

$$\begin{aligned} &= \sum_{i=1}^m \mathbb{E} \left[ \xi_i^2 \mathbb{E} \left[ (B_{t_{i+1}} - B_{t_i})^2 | \mathcal{F}_{t_i} \right] \right] \\ &\quad + 2 \sum_{i>j=1}^m \mathbb{E} \left[ \xi_i \xi_j \mathbb{E} \left[ (B_{t_{i+1}} - B_{t_i}) (B_{t_{j+1}} - B_{t_j}) | \mathcal{F}_{t_i} \right] \right] \end{aligned}$$

Siden  $j < i$  følger det at  $B_{t_{j+1}} - B_{t_j}$  er målbare m.h.p.  $\mathcal{F}_{t_i}$  for alle  $j$ . Fra egenskap (iii) til den Brownske bevegelsen har vi og at  $B_{t_{i+1}} - B_{t_i}$  er uavhengig  $\mathcal{F}_{t_i}$ . Dette medfører at vi får:

$$\begin{aligned} E \left[ \left( \int_0^T \phi(s, \omega) dB_s(\omega) \right)^2 \right] &= \sum_{i=1}^m \mathbb{E} \left[ \xi_i^2 \mathbb{E} \left[ (B_{t_{i+1}} - B_{t_i})^2 \right] \right] \\ &\quad + 2 \sum_{i>j=1}^m \mathbb{E} \left[ \xi_i \xi_j (B_{t_{j+1}} - B_{t_j}) \mathbb{E} \left[ (B_{t_{i+1}} - B_{t_i}) \right] \right] \end{aligned}$$





Vi skriver det om på vektor og matriseform med vektorene  $f : [0, T] \times \mathbb{R}^d \rightarrow \mathbb{R}^d$ ,  $B_t : \Omega \rightarrow \mathbb{R}^m$  og matrisen  $\sigma : [0, T] \times \mathbb{R}^d \rightarrow \mathbb{R}^d \times \mathbb{R}^m$  og får:

$$dX_t = \begin{bmatrix} dX_t^1 \\ dX_t^2 \\ \vdots \\ dX_t^d \end{bmatrix} = \begin{bmatrix} f_1(t, X_t)dt \\ f_2(t, X_t)dt \\ \vdots \\ f_d(t, X_t)dt \end{bmatrix} + \begin{bmatrix} \sigma_{1,1}(t, X_t) & \sigma_{1,2}(t, X_t) & \dots & \sigma_{1,m}(t, X_t) \\ \sigma_{2,1}(t, X_t) & \ddots & & \vdots \\ \vdots & & \ddots & \vdots \\ \sigma_{d,1}(t, X_t) & \dots & \dots & \sigma_{d,m}(t, X_t) \end{bmatrix} \begin{bmatrix} dB_t^1 \\ dB_t^2 \\ \vdots \\ dB_t^m \end{bmatrix}$$

som forenkles til:

$$dX_t = f(t, X_t)dt + \sigma(t, X_t)dB_t$$

Vi skal nå formulere den multidimensjonale Itô's formel.

**Lemma 1.3** (Itô's formel). *Vi antar nå at vi har en funksjon  $g : [0, T] \times \mathbb{R}^d \rightarrow \mathbb{R}^k$  slik at  $g \in C^{1,2}(\mathbb{R}^k)$ . Da er  $Y(t) = g(t, X_t) = (g_1(t, X_t), g_2(t, X_t), \dots, g_k(t, X_t))$  en ny Itô prosess. For hver komponent  $l \in \{1, 2, \dots, k\}$  av  $Y$  har vi:*

$$dY^l = \frac{\partial g_k}{\partial t}(t, X_t)dt + \sum_i^d \frac{\partial g_k}{\partial x_i}(t, X_t)dX^i + \frac{1}{2} \sum_{i,j}^d \frac{\partial^2 g_k}{\partial x_i \partial x_j}(t, X_t)dX^i dX^j$$

Vi skal se på de to siste leddene i Itô's formel.

$$dY^l = \frac{\partial g_k}{\partial t}(t, X_t)dt + I + II$$

I det første leddet (I) får vi:

$$\begin{aligned} I &= \sum_i^d \frac{\partial g_k}{\partial x_i}(t, X_t)dX^i = \sum_i^d \frac{\partial g_k}{\partial x_i}(t, X_t) \left( f_i(t, X_t)dt + \sum_{l=1}^m \sigma_{i,l}(t, X_t)dB_t^l \right) \\ &= \sum_i^d \frac{\partial g_k}{\partial x_i}(t, X_t)f_i(t, X_t)dt + \sum_i^d \sum_{l=1}^m \frac{\partial g_k}{\partial x_i}(t, X_t)\sigma_{i,l}(t, X_t)dB_t^l \end{aligned}$$

Fra ledd (II) får vi:

$$\begin{aligned} II &= \frac{1}{2} \sum_{i,j}^d \frac{\partial^2 g_k}{\partial x_i \partial x_j}(t, X_t)dX^i(dX^j) \\ &= \frac{1}{2} \sum_{i,j}^d \frac{\partial^2 g_k}{\partial x_i \partial x_j}(t, X_t) \left( (f_i(t, X_t)dt + \sum_{l=1}^m \sigma_{i,l}(t, X_t)dB_t^l) (f_j(t, X_t)dt + \sum_{q=1}^m \sigma_{j,q}(t, X_t)dB_t^q) \right) \\ &= \frac{1}{2} \sum_{i,j}^d \frac{\partial^2 g_k}{\partial x_i \partial x_j}(t, X_t) \left( f_i(t, X_t)f_j(t, X_t)(dt)^2 + f_i(t, X_t) \sum_{q=1}^m \sigma_{i,q}(t, X_t)dB_t^q dt \right. \\ &\quad \left. + \sum_{l=1}^m \sigma_{i,l}(t, X_t)f_j(t, X_t)dB_t^l dt + \left( \sum_{l=1}^m \sigma_{i,l}(t, X_t)dB_t^l \right) \left( \sum_{q=1}^m \sigma_{j,q}(t, X_t)dB_t^q \right) \right) \end{aligned}$$

For å forenkle uttrykket over trenger vi en multiplikasjonstabell hentet fra (Karatzas & Shreve, 1991, side 154) for å utregning av integraler. Dersom  $\overline{B}_t$  og  $B_t$  uavhengige av hverandre får vi :

Tabell 1:

$\cdot$	$dB_t$	$d\overline{B}_t$	$dt$
$dB_t$	$dt$	0	0
$d\overline{B}_t$	0	$dt$	0
$dt$	0	0	0

Ved bruk av denne tabellen forsvinner leddene med  $dt^2$  og  $dB_t^l dt$ . I siste leddet får vi fra symmetri:

$$\sum_{l=1}^m \sigma_{i,l}(t, X_t) dB_t^l \left( \sum_{l=1}^m \sigma_{i,l}(t, X_t) dB_t^l \right) = \sum_{l=1}^m \sigma_{i,l}^2(t, X_t) (dB_t^l)^2 + 2 \sum_{l>q=1}^m \sigma_{i,l}(t, X_t) \sigma_{i,q}(t, X_t) dB_t^l dB_t^q$$

og det følger fra tabellen at

$$\sum_{l>q=1}^m \sigma_{i,l}(t, X_t) \sigma_{i,q}(t, X_t) dB_t^l dB_t^q = 0$$

ettersom  $dB_t^l$  og  $dB_t^q$  er uavhengige av hverandre. Vi står da igjen med:

$$\begin{aligned} dY^l &= \frac{\partial g_k}{\partial t}(t, X_t) dt + \sum_i^d \frac{\partial g_k}{\partial x_i}(t, X_t) f_i(t, X_t) dt + \sum_i^d \sum_{l=1}^m \frac{\partial g_k}{\partial x_i}(t, X_t) \sigma_{i,l}(t, X_t) dB_t^l \\ &+ \frac{1}{2} \sum_{i,j}^d \sum_{l=1}^m \frac{\partial^2 g_k}{\partial x_i \partial x_j}(t, X_t) \sigma_{i,l}^2(t, X_t) dt \end{aligned}$$

Dersom  $X_t$  er en endimensjonal prosess og  $g : [0, T] \times \mathbb{R} \rightarrow \mathbb{R}$  får vi fra Itô's formel:

$$\begin{aligned} dY &= \frac{\partial g}{\partial t}(t, X_t) dt + \frac{\partial g}{\partial x}(t, X_t) dX_t + \frac{1}{2} \frac{\partial^2 g}{\partial x^2}(t, X_t) dX_t(dX_t) \\ &= \left[ \frac{\partial g}{\partial t}(t, X_t) + \frac{\partial g}{\partial x}(t, X_t) f(t, X_t) + \frac{1}{2} \frac{\partial^2 g}{\partial x^2}(t, X_t) \sigma(t, X_t)^2 \right] dt \\ &+ \frac{\partial g}{\partial x}(t, X_t) \sigma(t, X_t) dB_t \end{aligned}$$

### 1.2.3 Martingaler

La  $(\Omega, \mathcal{F})$  være gitt og  $\{\mathcal{F}_t\}_{t \geq 0}$  er en filtrasjon på  $\mathcal{F}$ . Da er en stokastiske prosess  $\{X_t\}_{t \geq 0}$  en martingal om den tilfredsstill:

**Definisjon 1.12** (Martingale-egenskapen). *La  $X_t$  være adaptert til filtrasjonen  $\mathcal{F}_t$  og  $\mathbb{E}[|X|] < \infty$ .*

$$\mathbb{E}[X_t | \mathcal{F}_s] = X_s, \quad \forall t \in [0, T] \quad (13)$$

En viktig klasse med martingaler er Itô-integralene. Vi står over beviset, men beviset bygger på at den Brownske bevegelsen er en martingal. At Itô integraler er martingaler gir oss:

$$\mathbb{E}\left[\int_0^T \phi(t, \omega) dB_t | \mathcal{F}_s\right] = \int_0^s \phi(t, \omega) dB_t$$

Som igjen, ved bruk av tårnegenskapen til den betingede forventningen, resulterer i at:

$$\mathbb{E}\left[\int_0^T \phi(t, \omega) dB_t\right] = \mathbb{E}\left[\mathbb{E}\left[\int_0^T \phi(t, \omega) dB_t | F_0\right]\right] = \mathbb{E}\left[\int_0^0 \phi(t, \omega) dB_t\right] = 0 \quad (14)$$

#### 1.2.4 Diffusjonsprosesser

Vi skal videre holde oss til en klasse SDE-er kalt tidshomogene Itô-diffusjoner.

**Definisjon 1.13** (Itô diffusjon(tidshomogen)). *Itô diffusjon er en stokastisk prosess  $X_t(\omega) = X(t, \omega) : [0, t) \times \Omega \rightarrow \mathbb{R}^n$  som tilfredstiller den stokastiske differensialligningen:*

$$dX_t = b(X_t)dt + \sigma(X_t)dB_t \quad t \geq s; \quad X_s = x \quad (15)$$

Vi deler diffusjonsprosesser inn i to ledd. Det første leddet  $b(X_t)$  er driften og det andre leddet  $\sigma(X_t)$  diffusjonen. Det at begge leddene er uavhengige av  $t$  medfører at diffusjonsprosesser besitter det vi kaller Markov-egenskapen.

**Definisjon 1.14** (Markov-egenskapen). *Gitt et sannsynlighetsrom  $(\Omega, \mathcal{F}, P)$  med en filtrasjon  $\{\mathcal{F}_t\}_{t \in [0, \infty)}$  og  $X_t : \Omega \rightarrow \mathbb{R}^n$ .*

*Vi sier at  $X_t$  har Markov-egenskapen dersom:*

$$E^x[f(X_{t+s}) | \mathcal{F}_t](\omega) = E^{X_t(\omega)}[f(X_s)]$$

$E^x$  refererer til den betingede forventningen gitt  $X_0 = x$ .

Markov-egenskapen forteller oss at gitt informasjonen  $\mathcal{F}_t$  om prosessen frem til  $t$ , er det kun informasjonen om den siste posisjonen  $X_t$  som er relevant. I eksempel 1.2, med kron og mynt, har vi en forventning etter 7 kast på 3.5. Gitt nå at vi vet at etter 3 dager har vi oppnådd 2. Da påvirker ikke rekkefølge vi har fått mynt og kron, de foregående dagene, forventningen det er kun hva vi har oppnådd etter dag 3 som påvirker forventningen. En stokastisk prosess med Markov-egenskaper kalles en Markov-prosess.

### 1.2.5 Ornstein-Uhlenbeck prosess

En Ornstein-Uhlenbeck prosess er en diffusjonsprosess med konstante koeffisienter  $\mu$ ,  $\sigma$  og  $\theta$ . Dette er en av få klasser med stokastiske prosesser med en analytisk løsning.

**Definisjon 1.15** (Ornstein-Uhlenbeck prosess). La  $\theta > 0$ ,  $\sigma > 0$  og  $\mu$  være konstanter. Da er løsningen  $X_t$  av:

$$\begin{cases} dX_t &= \theta(\mu - X_t)dt + \sigma dB_t & \theta \neq 0 \\ X_0 &= x_0 \end{cases} \quad (16)$$

en Ornstein-Uhlenbeck prosess.

Ved bruk av Itô's formel og et riktig valg av  $Y(t) = g(t, X_t)$  på en Ornstein-Uhlenbeck prosess, finner vi løsningen  $X_t$ .

$$\text{velg: } Y(t) = g(t, X_t) = X_t e^{\theta t}$$

Da følger:

$$\partial_t g(t, X_t) = \theta X_t e^{\theta t}, \quad \partial_X g(t, X_t) = e^{\theta t}, \quad \partial_{X^2}^2 g(t, X_t) = 0$$

Setter vi dette inn i Itô's formel får vi:

$$dY(t) = \theta X_t e^{\theta t} dt + e^{\theta t} (\theta(\mu - X_t) dt + \sigma dB_t)$$

Utrykket kan vi nå forenkles til:

$$dY(t) = e^{\theta t} \theta \mu dt + e^{\theta t} \sigma dB_t$$

Bytter  $Y(t)$  med  $X_t e^{\theta t}$  og integrerer med hensyn på  $t$  får vi:

$$X_t e^{\theta t} = x_0 + \int_0^t e^{\theta s} \theta \mu ds + \int_0^t e^{\theta s} \sigma dB_s$$

Vi finner nå løsningen  $X_t$  ved å multipliserer med  $e^{-\theta t}$ :

$$X_t = x_0 e^{-\theta t} + e^{-\theta t} \theta \mu \int_0^t e^{\theta s} ds + e^{-\theta t} \sigma \int_0^t e^{\theta s} dB_s$$

Ut fra løsningen kan vi finne forventningen og variansen til prosessen  $X_t$  og dermed den tilhørende fordelingen til prosessen:

**Lemma 1.4** (Fordelingen til Ornstein-Uhlenbeck prosesser).  $X_t$  er en Ornstein-Uhlenbeck prosess. Da  $X_t$  normalfordelt ( $\sim \mathcal{N}$ ) med fordelingen:

$$X_t \sim \mathcal{N}\left(X_0 e^{-\theta t} + \mu(1 - e^{-\theta t}), \frac{\sigma^2}{2\theta}(1 - e^{-2\theta t})\right)$$

Før vi beviser Lemma 1.4 ser vi at  $X_t$  består av to deler. En deterministisk del bestående av de første to leddene;  $x_0 e^{-\theta t} + e^{-\theta t} \theta \mu \int_0^t e^{\theta s} dt$  som kun er avhengig av  $t$ , og en del  $2 \sigma \int_0^t e^{-\theta(t-s)} dB_s$  som er et Itô-integral.

*Bevis Lemma 1.4.* Forventningen til Ornstein-Uhlenbeck prosessen,  $X_t$  er gitt ved den deterministiske delen av løsningen  $X_t$  siden Itô-integralet er en martingal med forventning 0, (14):

$$E[X_t] = x_0 e^{-\theta t} + \mu \int_0^t \theta e^{-\theta(t-s)} dt = x_0 e^{-\theta t} + \mu e^{-\theta t} \left[ \frac{\theta e^{\theta s}}{\theta} \right]_0^t = x_0 e^{-\theta t} + \mu e^{-\theta t} (e^{\theta t} - 1)$$

For å finne variansen bruker vi at  $Var(a + X) = Var(X)$  når  $a$  er deterministisk, så variansen til prosessen er kun avhengig av Itô-integralet. Ved bruk av Itô-isometrien lemma 1.2, finner vi variansen:

$$Var(X_t) = E\left[\left(\int_0^t e^{-\theta(t-s)} \sigma dB_s\right)^2\right] = E\left[\int_0^t (e^{-\theta(t-s)} \sigma)^2 ds\right] = \frac{\sigma^2}{2\theta} (1 - e^{-2\theta t})$$

Dette resulterer i fordelingen til en Ornstein-Uhlenbeck prosess:

$$X_t \sim \mathcal{N}\left(X_0 e^{-\theta t} + \mu(1 - e^{-\theta t}), \frac{\sigma^2}{2\theta}(1 - e^{-2\theta t})\right) \quad (17)$$

□

### 1.3 Sammenheng mellom PDE-er og SDE-er

Det finnes en sammenheng mellom diffusjonsprosesser og løsninger av partielle differensialligninger. Innenfor matematikken ble dette kjent som Kolmogorovs fremover og bakover ligninger. Fremoverligningen ble og oppdaget uavhengig innenfor fysikken noen år tidligere hvor det kalles Fokker-Planck-ligningen.

En stokastisk differensialligning er knyttet til Fokker-Planck-ligningen ved: Dersom det finnes en entydig løsning av den tilhørende Fokker-Planck-ligningen, er denne løsningen tettheten til diffusjonsprosessen  $X_t$  som løser den stokastiske differensialligningen. Før vi finner Fokker-Planck-ligningen trenger vi det som kalles generatoren til en diffusjonsprosess.

#### 1.3.1 Generatoren til diffusjoner

Vi starter med en  $d$ -dimensjonal diffusjon  $X_t$  med  $b : \mathbb{R}^d \rightarrow \mathbb{R}^d$  og  $\sigma : \mathbb{R}^d \rightarrow \mathbb{R}^{d \times m}$  og  $B_t$  er en  $m$ -dimensjonal Brownsk bevegelse:

$$\begin{cases} dX_t &= b(X_t)dt + \sigma(X_t)dB_t \\ X_0 &= x_0 \quad x_0 \in \mathbb{R}^d \end{cases} \quad (18)$$

Vi lar nå  $\mathcal{D}_A(X_t)$  være mengden funksjoner  $f : \mathbb{R}^d \rightarrow \mathbb{R}$  hvor grensen under holder:

$$A[f(X_t)] = \lim_{t \rightarrow 0} \frac{\mathbb{E}^x[f(X_t)] - f(x)}{t} \quad x \in \mathbb{R}^d$$

$\mathbb{E}^x$  er en kortform av  $\mathbb{E}[\cdot | X_0 = x]$  betyr her den betingede forventningen gitt  $X_0 = x$ . Hvis vi videre antar at  $f$  er tilstrekkelig glatt og med kompakt støtte, får vi generatoren til diffusjonen (18) ved:

**Teorem 1.5** (Generator til Itô-diffusjonsprosessen  $X_t$ ). *La  $X_t$  være en itô-diffusjonsprosess og  $f \in C_0^2(\mathbb{R}^d)$  da er generatoren  $A$  til  $X_t$ :*

$$A[f(x)] = \sum_{i=1}^d b_i(x) \frac{\partial f}{\partial x_i} + \frac{1}{2} \sum_{i,j} \sigma(x) \sigma(x)_{i,j}^T \frac{\partial^2 f}{\partial x_i \partial x_j} \quad (19)$$

*Bevis.* Anta at  $f : \mathbb{R}^d \rightarrow \mathbb{R}$  og siden  $f \in C_0^2(\mathbb{R}^d)$ . Da har vi fra den multidimensjonale Itô-formelen:

$$\begin{aligned} Y(t) &= f(X_t) \\ dY(t) &= df(X_t) = \sum_i \frac{\partial f}{\partial x_i}(X_t) dX_i + \frac{1}{2} \sum_{i,j} \frac{\partial^2 f}{\partial x_i \partial x_j}(X_t) (dX_i)(dX_j) \\ &= \sum_i \frac{\partial f}{\partial x_i}(X_t) (b_i(x) dt + \sigma_i(X_t) dB_t) \\ &\quad + \frac{1}{2} \sum_{i,j} \frac{\partial^2 f}{\partial x_i \partial x_j}(X_t) \left( (b_i(X_t) dt + \sigma_i(X_t) dB_t) \left( (b_j(X_t) dt + \sigma_j(X_t) dB_t) \right)^T \right) \end{aligned}$$

Hvor  $\sigma_i(X_t)$  er den  $i$ -te radvektoren til matrisen  $\sigma(X_t)$ .

Vi kan nå bruke multiplikasjonstabellen (1) og regneregler for vektorer for å forenkle uttrykket til:

$$= \sum_{i=1}^d \frac{\partial f}{\partial x_i}(X_t) (b_i(X_t) dt + \sigma_i(X_t) dB_t) + \frac{1}{2} \sum_{i,j} \frac{\partial^2 f}{\partial x_i \partial x_j}(X_t) \sigma_i(X_t) \sigma_j(X_t)^T dt$$

Etter integrasjon på begge sider m.h.p.  $t$  får vi:

$$\begin{aligned} f(X_t) &= f(x) + \int_0^t \left( \sum_{i=1}^d \frac{\partial f}{\partial x_i}(X_s) b_i(X_s) + \frac{1}{2} \sum_{i,j} \frac{\partial^2 f}{\partial x_i \partial x_j}(X_s) \sigma_i(X_s) \sigma_j(X_s)^T \right) ds \\ &\quad + \int_0^t \sum_i \sigma_i(X_s) dB_s \end{aligned}$$

Vi tar den betingede forventningen gitt  $X_0 = x$  på begge sider av likheten og bruker martingalegenskapen (14) til  $\int_0^t \sigma_i(X_s) dB_s$  og får:

$$\mathbb{E}^x[f(X_t)] = f(x) + \mathbb{E}^x \left[ \int_0^t \left( \sum_{i=1}^d \frac{\partial f}{\partial x_i}(X_s) b_i(X_s) + \frac{1}{2} \sum_{i,j} \frac{\partial^2 f}{\partial x_i \partial x_j}(X_s) \sigma_i(X_s) \sigma_j(X_s)^T \right) ds \right] \quad (20)$$

Så med litt manipulering av uttrykket over står vi igjen med:

$$\begin{aligned} \frac{\mathbb{E}^x[f(X_t)] - f(x)}{t} &= \frac{\mathbb{E}^x \left[ \int_0^t \left( \sum_{i=1}^d \frac{\partial f}{\partial x_i}(X_s) b_i(X_s) + \frac{1}{2} \sum_{i,j} \frac{\partial^2 f}{\partial x_i \partial x_j}(X_s) \sigma_i(X_s) \sigma_j(X_s)^T \right) ds \right]}{t} \\ &= \int_0^t \left( \frac{\mathbb{E}^x \left[ \sum_{i=1}^d \frac{\partial f}{\partial x_i}(X_s) b_i(X_s) + \frac{1}{2} \sum_{i,j} \frac{\partial^2 f}{\partial x_i \partial x_j}(X_s) \sigma_i(X_s) \sigma_j(X_s)^T \right]}{t} \right) ds \end{aligned}$$

Lar vi så  $t$  gå mot 0 får vi:

$$\sum_{i=1}^d \frac{\partial f}{\partial x_i}(x) b_i(x) + \frac{1}{2} \sum_{i,j} \frac{\partial^2 f}{\partial x_i \partial x_j}(x) \sigma_i(x) \sigma_j(x)^T = A[f(x)]$$

Som er det resultatet vi ønsket. □

### 1.3.2 Fokker-Planck-ligningen tilknyttet Itô-diffusjonsprosesser

Vi skal nå vise hvordan vi finner Fokker-Planck-ligningen til Itô diffusjonsprosesser.

**Lemma 1.6** (Fokker-Planck-ligningen). *Anta at  $X_t$  er en løsning av Itô-diffusjonsprosess (18). Da tilfredsstiller tettheten  $\rho(t, y; x)$  til  $X_t$ , hvor  $X_0 = x$ , Fokker-Planck-ligningen:*

$$\frac{\partial \rho}{\partial t}(t, y; x) = - \sum_{i=1}^d \frac{\partial}{\partial y_i} \left( b_i(y) \rho(t, y; x) \right) + \sum_{i,j} \frac{\partial^2}{\partial y_i \partial y_j} \left( D_{i,j}(y) \rho(t, y; x) \right)$$

$D_{i,j} = \frac{1}{2} \sigma_i \sigma_j^T$  hvor  $\sigma_i$  og  $\sigma_j$  er radvektorer til  $\sigma$ .

*Bevis.* For å finne Fokker-Planck-ligningen til en diffusjon (18) trenger vi  $f : \mathbb{R}^d \rightarrow \mathbb{R}$  og at  $f \in C^2(\mathbb{R})$ . Vi følger videre samme fremgangsmåten brukt over for å finne generatoren til diffusjonen, men stopper opp ved (20) :

$$\mathbb{E}^x[f(X_t)] = f(x) + \mathbb{E}^x \left[ \int_0^t \left( \sum_{i=1}^d \frac{\partial f}{\partial x_i}(X_s) b_i(X_s) + \frac{1}{2} \sum_{i,j} \frac{\partial^2 f}{\partial x_i \partial x_j}(X_s) \sigma_i(X_s) \sigma_j(X_s)^T \right) ds \right]$$



Vi antar nå at det finnes en tilhørende tetthetsfunksjon som er avhengig av  $X_0 = x$  og  $t$ . Da kan vi skrive forventningen:

$$\mathbb{E}^x[f(X_t)] = \int_{\Omega} f(X_t(\omega))dP(\omega) = \int_{\mathbb{R}^d} f(y)\rho(t, y; x)dy$$

Under denne antagelsen om at det finnes en tilhørende tetthet kan vi og skrive om høyre side av (20):

$$\begin{aligned} &= f(x) + \int_{\mathbb{R}^d} \left[ \int_0^t \left( \sum_{i=1}^d \frac{\partial f}{\partial x_i}(y)b_i(y) + \frac{1}{2} \sum_{i,j} \frac{\partial^2 f}{\partial x_i \partial x_j}(y)\sigma_i(y)\sigma_j(y)^T \right) ds \right] \rho(t, y; x)dy \\ &= f(x) + \int_0^t \left[ \int_{\mathbb{R}^d} \left( \sum_{i=1}^d \frac{\partial f}{\partial x_i}(y)b_i(y) + \frac{1}{2} \sum_{i,j} \frac{\partial^2 f}{\partial x_i \partial x_j}(y)\sigma_i(y)\sigma_j(y)^T \right) \rho(s, y; x)dy \right] ds \\ &= f(x) + \int_0^t \left[ \int_{\mathbb{R}^d} \left( \sum_{i=1}^d \frac{\partial f}{\partial x_i}(y)b_i(y) \right) \rho(s, y; x)dy \right] ds \\ &\quad + \int_0^t \left[ \int_{\mathbb{R}^d} \left( \frac{1}{2} \sum_{i,j} \frac{\partial^2 f}{\partial x_i \partial x_j}(y)\sigma_i(y)\sigma_j(y)^T \right) \rho(s, y; x)dy \right] ds = f(x) + I + II \end{aligned}$$

Før vi går videre og ser på  $I$  og  $II$  hver for seg må vi ha med et lite mineresultat:

**Lemma 1.7.** *la  $f$  og  $g$  være integrerbare på  $\mathbb{R}^d$ . Da følger:*

$$\int_{\mathbb{R}^d} \frac{\partial f}{\partial x_c}(x)g(x)dx = - \int_{\mathbb{R}^d} f(x)\frac{\partial g}{\partial x_c}(x)dx \quad (21)$$

*Bevis Lemma 1.7.* Vi tar utgangspunkt i produktregelen for derivasjon:

$$\frac{\partial}{\partial x_c}(f \cdot g)(x) = \frac{\partial f}{\partial x_c}(x)g(x) + f(x)\frac{\partial g}{\partial x_c}(x)$$

Integrerer vi på begge sider får vi:

$$\begin{aligned} \int_{\mathbb{R}^d} \frac{\partial f}{\partial x_c}(x)g(x)dx + \int_{\mathbb{R}^d} f(x)\frac{\partial g}{\partial x_c}(x)dx &= \int_{\mathbb{R}^d} \frac{\partial}{\partial x_c}(f \cdot g)(x)dx \\ &= \int_{\mathbb{R}^{d-1}} \left( \lim_{n \rightarrow \infty} [f(x_c, x)g(x_c, x)]_{-n}^n \right) dx = 0 \text{ når } n \rightarrow \infty \end{aligned}$$

Som gir oss ønsket resultat:

$$\int_{\mathbb{R}^d} \frac{\partial f}{\partial x_c}(x)g(x)dx = - \int_{\mathbb{R}^d} f(x)\frac{\partial g}{\partial x_c}(x)dx$$

□

Ser vi først på  $I$ :

$$\begin{aligned} I &= \int_0^t \left[ \int_{\mathbb{R}^d} \left( \sum_{i=1}^d \frac{\partial f}{\partial x_i}(y) b_i(y) \right) \rho(s, y; x) dy \right] ds \\ &= \int_0^t \left[ \sum_{i=1}^d \int_{\mathbb{R}^d} \frac{\partial f}{\partial x_i}(y) (b_i(y) \rho(s, y; x)) dy \right] ds \end{aligned}$$

Så bruker vi lemma 1.7 for å flytte den partiellderiverte fra  $\frac{\partial}{\partial x_i} f(y)$  over på  $b_i(y) \rho(s, y; x)$ .

$$\begin{aligned} I &= \int_0^t \left[ \sum_{i=1}^d \int_{\mathbb{R}^d} \frac{\partial f}{\partial x_i}(y) (b_i(y) \rho(s, y; x)) dy \right] ds \\ &= - \int_0^t \left[ \sum_{i=1}^d \int_{\mathbb{R}^d} f(y) \frac{\partial}{\partial x_i} (b_i(y) \rho(s, y; x)) dy \right] ds \end{aligned}$$

for  $II$  bruker vi teorem 1.7 2 ganger og får:

$$\begin{aligned} II &= \int_0^t \left[ \int_{\mathbb{R}^d} \left( \frac{1}{2} \sum_{i,j}^d \frac{\partial^2 f}{\partial x_i \partial x_j}(y) \sigma_i(y) \sigma_j(y)^T \right) \rho(s, y; x) dy \right] ds \\ &= \frac{1}{2} \int_0^t \left[ \sum_{i,j}^d \int_{\mathbb{R}^d} \frac{\partial^2 f}{\partial x_i \partial x_j}(y) \sigma_i(y) \sigma_j(y)^T \rho(s, y; x) dy \right] ds \\ &= \frac{1}{2} \int_0^t \left[ \sum_{i,j}^d \int_{\mathbb{R}^d} f(y) \frac{\partial^2}{\partial x_i \partial x_j} (\sigma_i(y) \sigma_j(y)^T \rho(s, y; x)) dy \right] ds \end{aligned}$$

så vi har kommet frem til:

$$\begin{aligned} \mathbb{E}^x[f(y)] &= f(x) - \int_0^t \left[ \sum_{i=1}^d \int_{\mathbb{R}^d} f(y) \frac{\partial}{\partial x_i} (b_i(y) \rho(s, y; x)) dy \right] ds \\ &\quad + \frac{1}{2} \int_0^t \left[ \sum_{i,j}^d \int_{\mathbb{R}^d} f(y) \frac{\partial^2}{\partial x_i \partial x_j} (\sigma_i(y) \sigma_j(y)^T \rho(s, y; x)) dy \right] ds \end{aligned}$$

Vi deriverer med hensyn på  $t$  og flytter summene innefor integralene igjen:

$$\begin{aligned} \frac{\partial}{\partial t} \mathbb{E}^x[f(X_t)] &= \frac{\partial}{\partial t} \left[ f(x) - \int_0^t \left[ \int_{\mathbb{R}^d} f(y) \sum_{i=1}^d \frac{\partial}{\partial x_i} (b_i(y) \rho(s, y; x)) dy \right] ds \right. \\ &\quad \left. + \frac{1}{2} \int_0^t \left[ \int_{\mathbb{R}^d} f(y) \sum_{i,j}^d \frac{\partial^2}{\partial x_i \partial x_j} (\sigma_i(y) \sigma_j(y)^T \rho(s, y; x)) dy \right] ds \right] \end{aligned}$$

$$\begin{aligned}
&= - \int_{\mathbb{R}^d} f(y) \sum_{i=1}^d \frac{\partial}{\partial x_i} (b_i(y) \rho(s, y; x)) dy + \frac{1}{2} \int_{\mathbb{R}^d} f(y) \sum_{i,j} \frac{\partial^2}{\partial x_i \partial x_j} (\sigma_i(y) \sigma_j(y)^T \rho(s, y; x)) dy \\
&= \int_{\mathbb{R}^d} \left[ - f(y) \sum_{i=1}^d \frac{\partial}{\partial x_i} (b_i(y) \rho(s, y; x)) + \frac{1}{2} f(y) \sum_{i,j} \frac{\partial^2}{\partial x_i \partial x_j} (\sigma_i(y) \sigma_j(y)^T \rho(s, y; x)) \right] dy \\
&= \int_{\mathbb{R}^d} f(y) \left[ - \sum_{i=1}^d \frac{\partial}{\partial x_i} (b_i(y) \rho(s, y; x)) + \frac{1}{2} \sum_{i,j} \frac{\partial^2}{\partial x_i \partial x_j} (\sigma_i(y) \sigma_j(y)^T \rho(s, y; x)) \right] dy
\end{aligned}$$

Fra antagelsen om at det finnes en tetthet  $\rho(t, y; x)$  får vi på venstre side :

$$\frac{\partial}{\partial t} \mathbb{E}^x [f(X_t)] = \frac{\partial}{\partial t} \left( \int_{\mathbb{R}^d} f(y) \rho(t, y; x) dy \right) = \int_{\mathbb{R}^d} f(y) \frac{\partial \rho}{\partial t}(t, y; x) dy$$

Som gjør at vi igjen har:

$$\int_{\mathbb{R}^d} f(y) \frac{\partial \rho}{\partial t}(t, y; x) dy = \int_{\mathbb{R}^d} f(y) \left[ - \sum_{i=1}^d \frac{\partial}{\partial x_i} (b_i(y) \rho(t, y; x)) + \frac{1}{2} \sum_{i,j} \frac{\partial^2}{\partial x_i \partial x_j} (\sigma_i(y) \sigma_j(y)^T \rho(t, y; x)) \right] dy \quad (22)$$

Siden dette holder for alle  $f \in C^2$  har vi:

$$\frac{\partial \rho}{\partial t}(t, y; x) = - \sum_{i=1}^d \frac{\partial}{\partial y_i} (b_i(y) \rho(t, y; x)) + \frac{1}{2} \sum_{i,j} \frac{\partial^2}{\partial y_i \partial y_j} (\sigma_i(y) \sigma_j(y)^T \rho(t, y; x)) \quad (23)$$

Dette er Fokker-Planck-ligningen tilhørende diffusjonen (18). Vi kan forekle ligningen ved å sette:

$$D = \frac{1}{2} \sigma \sigma^T \quad (24)$$

Og da er:

$$D_{i,j} = \frac{1}{2} \sum_{k=1}^n \sigma_{i,k} \sigma_{k,j} = \frac{1}{2} \sigma_i \sigma_j^T$$

Hvor  $\sigma_i$  er i-te radvektor til matrisen  $\sigma$  og vi kan skrive om (23) slik at vi får Fokker-Planck-ligningen til diffusjonen(18):

$$\frac{\partial \rho}{\partial t}(y; x, t) = - \sum_{i=1}^d \frac{\partial}{\partial y_i} (b_i(y) \rho(y; x, t)) + \sum_{i,j} \frac{\partial^2}{\partial y_i \partial y_j} (D_{i,j}(y) \rho(y; x, t))$$

□

Dette resultatet viser er at dersom vi løser Fokker-Planck-ligningen tilhørende  $X_t$ , finner vi tettheten  $\rho(t, y; x)$  til  $X_t$ . Vi skal vise dette med et enkelt eksempel: La  $\theta > 0$  og  $\sigma > 0$ .

$$dX_t = -\theta dt + \sigma dB_t, \quad X_0 = x_0 \quad t > 0 \quad (25)$$

Denne prosessen har løsningen:

$$X_t = x_0 - \theta t + \sigma B_t$$

Og fordelingen til løsningen  $X_t \sim \mathcal{N}(x_0 - \theta t, \sigma^2 t)$  med den tilhørende tettheten:

$$\frac{1}{\sqrt{2\pi\sigma^2 t}} \exp\left(\frac{-(x - x_0 + \theta t)^2}{2\sigma^2 t}\right)$$

Vi skal nå løse Fokker-Plank-ligningen, men før vi gjør det trenger vi et par resultater fra Fourier analyse. Resultatene under hentet fra (Bahouri et al., 2011, kapittel 1) for å løse dette problemet.

**Definisjon 1.16** (Fourier transform). *La  $f$  være integrerbar. Da er Fourier transformen  $\mathcal{F}[\cdot]$  til  $f$  definert:*

$$\mathcal{F}[f(x)](\xi) = \int_{-\infty}^{\infty} f(x)e^{-ix \cdot \xi} dx$$

Vi skriver:

$$\mathcal{F}[f(x)](\xi) = \widehat{f}(\xi)$$

$x \cdot \xi$  er prikkproduktet mellom  $x$  og  $\xi$ .

Fra definisjonen får vi:

**Proposisjon 1.1.** *La  $f$  være en integrerbar funksjon. Da holder:*

$$(1) \quad \mathcal{F}[f(x+h)](\xi) = \widehat{f}(\xi)e^{ih \cdot \xi}$$

$$(2) \quad \mathcal{F}\left[\frac{d}{dx}f(x)\right](\xi) = i\xi\widehat{f}(\xi)$$

*Bevis.* Vi starter med å bevise(1):

$$\mathcal{F}[f(x+h)](\xi) = \int_{-\infty}^{\infty} f(x+h)e^{-ix \cdot \xi} dx = \lim_{n \rightarrow \infty} \int_{-n}^n f(x+h)e^{-ix \cdot \xi} dx$$

Variabelbytte  $y = x + h$  gir  $x = y - h$  og  $dy = dx$ :

$$\lim_{n \rightarrow \infty} \int_{-n+h}^{n-h} f(y)e^{-i(y-h) \cdot \xi} dy = \int_{-\infty}^{\infty} f(y)e^{ih \cdot \xi} e^{-iy \cdot \xi} dy = \widehat{f}(\xi)e^{ih \cdot \xi}$$

Litt forenklet bevis av (2) i  $\mathbb{R}$ :

$$\begin{aligned} \mathcal{F}\left[\frac{d}{dx}f(x)\right](\xi) &= \mathcal{F}\left[\lim_{h \rightarrow 0} \frac{f(x+h) - f(x)}{h}\right](\xi) = \lim_{h \rightarrow 0} \frac{1}{h} (\mathcal{F}[f(x+h) - f(x)]) \\ &= \lim_{h \rightarrow 0} \frac{1}{h} (\mathcal{F}[f(x+h)] - \mathcal{F}[f(x)]) = \lim_{h \rightarrow 0} \frac{\widehat{f}(\xi)e^{ih\xi} - \widehat{f}(\xi)}{h} \\ &= \widehat{f}(\xi) \lim_{h \rightarrow 0} \frac{e^{ih\xi} - e^{i0\xi}}{h - 0} = \widehat{f}(\xi) \frac{d}{dh} e^{ih\xi} \Big|_{h=0} = i\xi\widehat{f}(\xi) \end{aligned}$$

□

**Definisjon 1.17** (Konvolusjon). *La  $f$  og  $g$  være to integrerbare funksjoner. Da er konvolusjonen mellom  $f$  og  $g$  definert:*

$$(f * g)(x) = \int f(y)g(x - y)dy \quad (26)$$

Videre har vi Fourier transformen til konvolusjonen mellom to funksjoner  $f$  og  $g$ :

**Proposisjon 1.2.** *La  $f$  og  $g$  være integrerbare. Da er Fourier transformen til konvolusjonen mellom  $f$  og  $g$ :*

$$\mathcal{F}[f * g(x)](\xi) = \widehat{f}(\xi)\widehat{g}(\xi) \quad (27)$$

*Bevis.* Vi går bare gjennom beviset for  $x \in \mathbb{R}$ .

$$\begin{aligned} \mathcal{F}[f * g(x)](\xi) &= \int_{-\infty}^{\infty} (f * g)(x)e^{-ix\xi} dx = \int_{-\infty}^{\infty} \left( \int_{-\infty}^{\infty} f(y)g(x - y)dy \right) e^{-ix\xi} dx \\ &= \int_{-\infty}^{\infty} f(y) \left( \int_{-\infty}^{\infty} g(x - y)e^{-ix\xi} dx \right) dy = \int_{-\infty}^{\infty} f(y) \left( \int_{-\infty}^{\infty} g(z)e^{-i(z+y)\xi} dz \right) dy \\ &= \int_{-\infty}^{\infty} f(y)e^{-iy\xi} \left( \int_{-\infty}^{\infty} g(z)e^{-iz\xi} dz \right) dy = \widehat{g}(\xi) \int_{-\infty}^{\infty} f(y)e^{-iy\xi} dy = \widehat{g}(\xi)\widehat{f}(\xi) \end{aligned}$$

(1) Bruker variabelskifte  $z = x - y$ . □

Det siste resultatet vi trenger for å løse Fokker-Planck-ligningen er knyttet til Gaussiske kjerner Vi har bare bruk for tilfellet hvor  $x \in \mathbb{R}$ .

**Definisjon 1.18** (Gaussisk kerne). *la  $\delta > 0$  og  $x \in \mathbb{R}$ . Da er en Gaussisk kerne  $G_{\delta}(\cdot)$ :*

$$G_{\delta}(x) = \frac{1}{\sqrt{2\pi\delta}} e^{-\frac{x^2}{2\delta}}$$

**Proposisjon 1.3** (Fourier transformen av en Gaussisk kerne).

$$\widehat{G_{\delta}(x)}(\xi) = e^{-\delta\frac{\xi^2}{2}}$$

Vi hopper over beviset da dette er litt langt og ikke gir mer nødvendig innsikt.

Vi har nå det vi trenger for å vise at løsningen  $\rho(t, x)$  til Fokker-Planck-ligningen er tettheten til  $X_t$  fra (25):

**Proposisjon 1.4.** *Løsningen  $\rho(t, x)$  av til Fokker-Planck-ligningen til (25) er tettheten til (25):*

$$\rho(t, x) = \frac{1}{\sqrt{2\pi\sigma^2 t}} e^{-\frac{(x-x_0+\theta t)^2}{2\sigma^2 t}}$$

*Bevis.* Fra (25) ser vi at  $b = \theta$  og  $D = \frac{1}{2}\sigma^2$  som gir en Fokker-Planck ligning:

$$\begin{cases} \frac{\partial}{\partial t}\rho(t, x) &= \theta \frac{\partial}{\partial x}\rho(t, x) + \frac{1}{2}\sigma^2 \frac{\partial^2}{\partial x^2}\rho(t, x) \\ \rho(0, x) &= \delta(x_0) \end{cases}$$

$\rho(0, x) = \delta(x_0)$  er dirac-delta ettersom  $P(X_0 = x_0) = 1$ . Denne ligningen kan vi løse ved bruk av Fourier-transformen og proposisjon 1.1:

$$\frac{\partial}{\partial t}\widehat{\rho}(t, x) = \theta i\xi\widehat{\rho} - \xi^2 \frac{1}{2}\sigma^2\widehat{\rho} \quad , \quad \widehat{\rho(0, x)} = \widehat{\delta(x_0)}$$

Dette er en ODE vi kan løse og får:

$$\widehat{\rho} = \widehat{\delta(x_0)}e^{(-\theta i\xi + \xi^2 \frac{1}{2}\sigma^2)t}$$

Vi må nå finne tilbake fra  $\widehat{\rho}$  til  $\rho$ :

$$\begin{aligned} \mathcal{F}[\rho(t, x)](\xi) &= \widehat{\rho}(t, \xi) = \widehat{\delta(x_0)}e^{(-\theta i\xi + \xi^2 \frac{1}{2}\sigma^2)t} = \widehat{\delta(x_0)}e^{i\theta t}e^{-\sigma^2 t \frac{\xi^2}{2}} \\ &\stackrel{1}{=} \widehat{\delta(x_0)}e^{i\theta t}\widehat{G_{\sigma^2 t}}(x) \\ &\stackrel{2}{=} \widehat{\delta(x_0)}\widehat{G_{\sigma^2 t}}(x + \theta t) \\ &\stackrel{3}{=} \mathcal{F}[\delta(x_0) * G_{\sigma^2 t}(x + \theta t)](\xi) \quad \implies \\ \rho(t, x) &= \delta(x_0) * \frac{1}{\sqrt{2\pi\sigma^2 t}}e^{-\frac{(x+\theta t)^2}{2\sigma^2 t}} \\ &= \int \delta(x_0)(y) \frac{1}{\sqrt{2\pi\sigma^2 t}}e^{-\frac{(x-y+\theta t)^2}{2\sigma^2 t}} dy \\ &= \frac{1}{\sqrt{2\pi\sigma^2 t}}e^{-\frac{(x-x_0+\theta t)^2}{2\sigma^2 t}} \end{aligned}$$

Etter litt manipulasjon av uttrykket kan vi i første overgang (1) bruker at siste faktor  $\exp(-\sigma^2 t \frac{\xi^2}{2})$  er Fourier transformen av en Gaussisk kjerne (1.18) med  $\delta = \sigma^2 t$ . I andre overgang (2) bruker vi proposisjon 1.1. Før vi i (3) bruker proposisjon 1.2, baklengs, at Fourier transformen til en konvolusjon er produktet av Fourier transformene til funksjonene.

Denne løsningen av Fokker-Planck-ligningen er den sammen som fordelingen til diffusjonen som var det vi ønsket å vise.

□

## 2 Optimal kontrollteori

Teorien i denne delen er i hovedsak hentet fra Evans enkle fremstilling av stoffet (Evans, u.å, kapittel 4 og 5) samt noe fra klassikeren til Fleming og Rishel, (Fleming, & Rishel, 1975, kapittel: 1,2).

Optimal kontrollteori er i dag et stort felt og det er mye utelatt i denne fremstillingen, men det er gjort et valg om å inkludere en gjennomgang av Pontryagins maksimalprinsipp i tillegg til metoden hvor man bruker Hamilton-Jacobiligningen. Dette for å gi et lite innblikk og forståelse av optimal kontrollteori.

Optimal kontrollteori har sine historiske røtter i variasjonskalkulusen utviklet for rundt 350 år siden. Denne retningen i matematikken ble mindre og mindre aktuell på begynnelsen av 1900-tallet, men rundt 1950 våknet den til liv i en ny form. Det utviklet seg et behov for å kontrollere dynamiske systemer bl.a. knyttet til ønsket om å effektivisere drivstoffbruken på fly. Løsningen på disse problemene fant man ved å bruke en generalisering av variasjonskalkulusen. Det er dette vi i dag kaller optimal kontrollteori. (Fleming, & Rishel, 1975, side 20)

Ideen bak teorien går ut på å kontrollere et dynamisk system med en form for kontroll  $\alpha(\cdot)$ :

$$ODE = \begin{cases} \dot{x}(t) = f(x(t), \alpha(t)) & t \in [0, \infty) \\ x(0) = x_0 \end{cases} \quad (28)$$

Vi kan tenke oss  $\alpha(\cdot)$  som en bryter eller en gasspedal vi kan trykke på for å endre systemet underveis. Vi vil at denne endringer gjør at systemet vi kontrollerer går i riktig retning. Vi kan f.eks. tenke oss en månerakett som faller mot bakken. Hvordan kan vi styre gassen på motoren slik at raketten lander rolig på bakken? Eller et system hvor man styrer produksjonen i et vannkraftanlegg. Da trenger vi kanskje en kontrollstrategi på produksjonen slik at vi maksimerer strømproduksjonen samtidig som det tas hensyn til miljøet rundt kraftverket.

Det er mange måter vi kan styre disse systemene på og vi trenger derfor en måte å måle hvordan resultatet av en gitt kontroll er opp mot hva vi ønsker. I kontrollteorien gjøres dette ved at vi innfører en avkastningsfunksjonal  $P[\cdot]$ . Hvor  $P[\cdot]$  er avhengig av hvilken kontroll  $\alpha(\cdot)$  vi bruker:

$$P[\alpha(\cdot)] := \int_0^\tau r(x(t), \alpha(t)) dt + g(x(\tau))$$

Avkastningsfunksjonen  $P$  måler altså hvor bra den gitte kontrollen på systemet er mot ønsket resultat. Oppgaven er å finne den kontrollen som fungerer slik at vi maksimerer avkastningsfunksjonen.

$$\max_{\alpha(\cdot)} P[\alpha(\cdot)] := \max_{\alpha(\cdot)} \int_0^\tau r(x(t), \alpha(t)) dt + g(x(\tau))$$

Vi kan dele funksjonen inn i to deler. Det første leddet består av et integral som vi kaller den løpende avkastning, mens siste ledd er terminal avkastning. Dersom det viktigste er å bare lande raketten vil den løpende avkastningen være 0. Det kun er posisjonen og farten til raketten ved den siste tiden  $\tau$  vi er

opptatt av. Dersom vi tenker oss at raketten skal ha en rolig oppbremsing vil vi måtte konstruere en løpende avkastning som ivaretar dette.

Vi skal se på bruk av både Pontryagins maksimalprinsipp og Hamilton-Jacobiligningen for å løse kontrollproblemer her og starter med førstnevnte.

## 2.1 Pontryagins maksimalprinsipp

Når vi skal bruke Pontryagins maksimalprinsipp finnes en rekke ulike problemer og tilhørende løsningsmetoder. Vi fokuserer på tilfellet hvor vi har fiksert ett siste tidspunkt  $T$  med et fritt endepunkt  $x(T)$ . Dette for å få en intuisjon av måten maksimalprinsippet brukes. Vi starter med en typisk problemstilling:

### 2.1.1 Optimal investeringsstrategi for $t \in [0, T]$

Vi har en bedrift hvor vi til en hver tid  $t$  har produserer verdier  $x(t)$ . Vi ønsker å maksimere hvor mye penger vi sitter igjen med ved en gitt endetid  $T$ . Vi kan i bedriften vår velge om vi vil investere verdier vi produserer tilbake i selskapet eller om vi skal ta dem ut. Andelen vi reinvesterer ved tiden  $t$  er gitt ved  $\alpha(t)x(t)$  hvor  $\alpha : [0, T] \rightarrow [0, 1]$ . La  $k > 0$  være en fiksert vekstrate på investeringene. Vi har da et dynamisk system hvor endringen av systemet er avhengig av valgte  $\alpha(\cdot)$ :

$$\begin{aligned} \dot{x}(t) &= k\alpha(t)x(t) & t \in (0, T] \\ x(0) &= x_0 & x_0 > 0 \end{aligned}$$

Vi ønsker å finne en optimal investeringsstrategi  $\alpha^*(t)$  slik at vi ved tiden  $T$  har maksimert avkastningen:

$$P[\alpha^*(\cdot)] = \max_{\alpha \in \mathcal{A}} \int_0^T (1 - \alpha(t))x(t)dt + 0$$

$\mathcal{A}$  er mengden med alle mulige kontrollstrategier  $\alpha(\cdot)$ . Vi har nå et problem med fiksert endetid  $T$ , med fritt endepunkt  $x(T)$ .

### 2.1.2 Pontryagins maksimalprinsipp for fiksert $T$

Pontryagins maksimalprinsipp har likheter med Lagrange multiplikatoren vi bruker når vi skal løse optimeringsproblemer med flere begrensninger. I stedet for å bruke  $\lambda$  har vi en funksjon  $p(\cdot)$  kalt costate som skal hjelpe oss å finne den optimale kontrollen  $\alpha^*(\cdot)$ .

La  $A \subseteq \mathbb{R}^m$  og  $f : \mathbb{R}^n \times A \rightarrow \mathbb{R}^n$  og  $x_0 \in \mathbb{R}^n$ . Da kan vi definere mengden med brukbare kontroller:

$$\mathcal{A} = \{\alpha(\cdot) : [0, \infty) \rightarrow A \mid \alpha(\cdot) \text{ er målbar}\} \quad (29)$$

Vi bygger videre fra den Hamiltonske funksjonen fra variasjonskalkulusen<sup>2</sup>:

<sup>2</sup>Utleddningen av den Hamiltonske funksjonen er å finne i bl.a. Evans notater.



**Definisjon 2.1** (Hamiltonske funksjonen). La  $x, p \in \mathbb{R}^n$  og  $a \in A$ . Da er den Hamiltonske funksjonen for kontrollteori:

$$H(x, p, a) := f(x, a)p + r(x, a)$$

Pontryagins maksimalprinsipp sier at dersom vi antar at  $\alpha^*(\cdot) \in \mathcal{A}$  er optimal for det dynamiske systemet og den tilhørende avkastningsfunksjonalen, og det finnes en tilhørende løsning av systemet  $x^*(\cdot)$ . Da finnes  $p^* : [0, T] \rightarrow \mathbb{R}^n$  slik at:

$$\dot{x}^*(t) = \nabla_p H(x^*(t), p^*(t), \alpha^*(t)) \quad (30)$$

$$\dot{p}^*(t) = -\nabla_x H(x^*(t), p^*(t), \alpha^*(t)) \quad (31)$$

$$p^*(T) = g(x^*(T)) \quad (32)$$

Videre følger at:

$$H(x^*(t), p^*(t), \alpha^*(t)) = \max_{\alpha \in \mathcal{A}} H(x(t), p(t), \alpha(t)) \quad t \in [0, T] \quad (33)$$

Hvor H er den Hamiltonske funksjonen 2.1.

### 2.1.3 Løsning for den optimale investeringsstrategi

Vi har  $A = [0, 1]$  og  $f : \mathbb{R} \times A \rightarrow \mathbb{R}$ . La nå  $\alpha(\cdot) = a$  en konstant.  $f(x, a) = kax$  og fra avkastningsfunksjonalen har vi  $r(x, a) = (1 - a)x$ . Dette gir oss den Hamiltonske funksjonen:

$$\begin{aligned} H(x, p, a) &= k \cdot a \cdot x \cdot p + (1 - a)x = x(k \cdot a \cdot p + 1 - a) \\ \dot{x}(t) &= H_p(x(t), p, \alpha(t)) = k \cdot \alpha(t) \cdot x(t) \\ \dot{p}(t) &= -H_x(x(t), p, \alpha(t)) = -k \cdot \alpha(t) \cdot p(t) - 1 + \alpha(t) \end{aligned}$$

Fra (33) skal vi maksimere  $H(x, p, a)$  for alle  $t \in [0, T]$  og ettersom  $x(t) > 0$  for alle t kan vi forenkle problemet til:

$$\max_{a \in A} H(x, p, a) = \max_{a \in A} (k \cdot a \cdot x \cdot p + (1 - a)x) = x(1 + \max_{a \in A} [a(k \cdot p - 1)])$$

Det holder å finne  $\max_{a \in A} [a(k \cdot p - 1)]$  for hver t.

$(kp - 1) \leq 0$  når  $p \leq 1/k$  siden k er en positiv konstant. Og ettersom  $a \in A = [0, 1]$  oppnår vi maksimum ved å velge  $a = 0$ . Når  $p > 1/k$  er  $(kp - 1) > 0$  og vi velger  $a = 1$ .

Når vi nå skal avgjøre for hvilke  $t \in [0, T]$  vi skal velge 0 og hvilke vi skal velge 1 trenger vi nå uttrykk for  $p(t)$ .

Fra  $P[\alpha(\cdot)]$  ser vi at  $g(x(T)) = 0$  så fra (32) har vi  $p(T) = 0$ . Vi har og antatt at  $p(t)$  er kontinuerlig og siden  $p(T) = 0$ , medfører kontinuiteten, at når t er nær nok T vil  $p(t) < 1/k$  og vi velger  $a = 0$ . Dette gir  $\dot{p}(t) = -1$ .

$$\int_t^T \dot{p}(s) ds = \int_t^T -1 ds \rightarrow p(T) - p(t) = -T + t \implies p(t) = T - t$$

Antar vi nå at  $T$  er stor nok finnes det en  $t^*$  slik at  $p(t^*) = 1/k$ . Da holder for  $t < t^* \implies p(t) > 1/k$ , og  $a = 1$ . Da blir  $\dot{p}(t) = -kp(t)$  med løsningen:

$$\begin{aligned} \dot{p}(t) + kp(t) &= 0, && \text{bruk integrerende faktor} \\ \int_t^{t^*} \left[ \left( e^{ks} p(s) \right)' \right] ds &= e^{kt^*} p(t^*) - e^{kt} p(t) \implies \\ p(t) &= \frac{1}{k} e^{k(t^*-t)} = \frac{1}{k} e^{k(T-\frac{1}{k}-t)} \quad t \in [0, T - \frac{1}{k}] \end{aligned}$$

Vi kan nå definere strategien:

$$\alpha(t) = \begin{cases} 1 & \text{Hvis: } t \in [0, T - 1/k) \\ 0 & \text{Hvis: } t \in [T - 1/k, T] \end{cases} \quad (34)$$

Med denne valgte strategien får vi det dynamiske systemet med  $x(0) = x_0$ :

$$\begin{aligned} \dot{x}(t) &= kx(t) \quad t \in [0, T - \frac{1}{k}) \\ \dot{x}(t) &= 0 \quad t \in [T - \frac{1}{k}, T] \end{aligned}$$

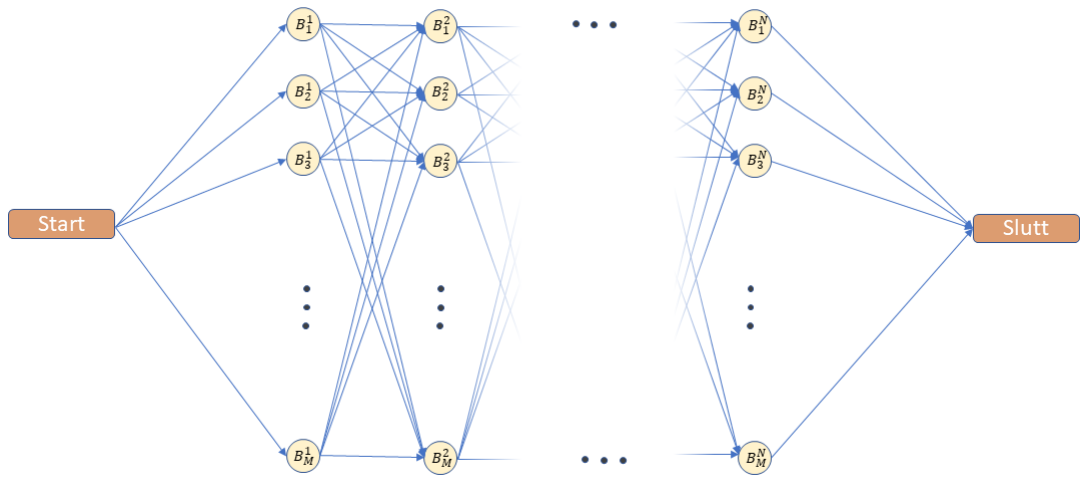
Og den maksimale avkastningen med kontroll  $\alpha(\cdot)$  blir:

$$\int_{T-1/k}^T x(t) dt = \frac{1}{k} x_0 e^{k(T-1/k)} = \frac{1}{k} x_0 e^{kT-1}$$

#### 2.1.4 Dynamisk programmering

Før vi tar for oss Hamilton-Jacobi-ligningen må vi først introdusere begrepet dynamisk programmering utviklet av Richard Bellman på 1950-tallet. Dynamisk programmering handler om å finne en effektiv måte å løse komplekse problemer. Ideen til Bellman var å løse en familie med tilsvarende komplekse problemer som inneholder det opprinnelige problemet som et delproblem. (Bellman, 1957, side 3) Dette virker paradoksalt, men det er rekkefølgen man løser alle delproblemene på som gjør at denne fremgangsmåten fungerer så bra.

**Eksempel 2.1.** *Tenk deg at du skal finne korteste vei mellom to byer. Du blir presentert med et forenklet kart i form av en graf med  $N$  lag. I hvert lag befinner det seg  $M$  byer og fra hver by kan du kjøre til alle byer i neste lag som illustrert her.*



Figur 2: Kart over mulige veier mellom Start og Mål

La  $d(\cdot, \cdot)$  være avstand mellom to byer i grafen og  $R_i^{l,m}$  er en unik rute i vi kan velge mellom by  $m$  i lag  $l$  og Slutt. Byene noteres med  $B_l^m$ . Vi vil finne den minste avstand  $\bar{d}(\text{start}, \text{slutt})$ . Dette er minimum av alle mulige ruter  $R_i^{0,1}$  mellom Start og Slutt.

$$\bar{d}(\text{Start}, \text{Slutt}) = \min\{R_1^{0,1}, R_2^{0,1}, \dots, R_r^{0,1}\}$$

Vi trenger derfor å finne alle  $r$  ruter vi kan velge. Fra Start har vi  $M$  mulige byer vi kan dra til i lag 1. Videre fra lag 1 har vi for hver by,  $M$  muligheter. Det vil si at vi har  $M \cdot M$  mulige ruter til lag 2. Slik fortsetter det helt til lag  $N$ . Fra lag  $N$  til Slutt er det bare en mulig rute så totalt antall ruter  $R_i^{0,1}$  vi kan velge mellom Start og Slutt er  $M^N$ . For hver rute vi velger er det  $N$  strekninger vi legger sammen så det gir antall regneoperasjoner  $N \cdot M^N$ .

Vi skal nå bruke prinsippet i dynamisk programmering og løse et større problem som inneholdene det opprinnelige problemet som et delproblem. Vi skal finne  $\bar{d}(B_l^m, \text{Slutt})$  for alle  $l \in (0, \dots, N)$  og  $m \in (1, 2 \dots M)$ . Altså løse problemet for alle byer inklusive Start.

Vi starter i lag  $N$ . Her er det bare en rute til Slutt for hver by. Går vi videre til lag  $(N-1)$  har hver by  $B_{N-1}^m$ ,  $M$  mulige ruter  $R_i^{(N-1),m}$  og

$$\bar{d}(B_{N-1}^m, \text{Slutt}) = \min\{R_1^{(N-1),m}, \dots, R_M^{(N-1),m}\}$$

Hver rute  $R_i^{(N-1),m} = d(B_{N-1}^m, B_N^i) + d(B_N^i, \text{Slutt})$  det betyr at for løse problemet for alle byer i lag  $(N-1)$  trenger vi  $M^2$  ruter, og for hver rute legger vi sammen to strekninger.

Videre i lag  $(N-2)$  har vi for hver av de  $M$  byene  $B_{N-2}^m$  rutene:

$$R_i^{(N-2),m} = d(B_{N-2}^m, B_{N-1}^i) + \bar{d}(B_{N-1}^i, \text{Slutt})$$

og for å løse korteste vei fra by  $B_{N-2}^m$  til Slutt finner vi:

$$\bar{d}(B_{N-2}^m, \text{Slutt}) = \min\{R_1^{(N-2),m}, \dots, R_M^{(N-2),m}\}$$

Så vi får tilsvarende som i lag  $(N-1)$ ,  $M^2$  ruter og for hver rute legges det sammen 2 strekninger.

Fortsetter vi slik helt til lag 1 vil få  $(N-1)M^2$  ruter vi må regne ut for å løse  $\bar{d}(B_l^m, \text{Slutt})$  for alle  $m \in (1, \dots, M)$  og  $l \in (1, \dots, N)$ .

Nå mangler vi bare det opprinnelige problemet med å finne:

$$\bar{d}(\text{Start}, \text{Slutt}) = \min\{d(\text{start}, B_1^1) + \bar{d}(B_1^1, \text{Slutt}), \dots, d(\text{start}, B_1^M) + \bar{d}(B_1^M, \text{Slutt})\}$$

Det er nå bare  $M$  mulige ruter ettersom vi allerede har regnet ut korteste vei fra lag 1. Hvis vi nå legger sammen får vi  $(N-1)M^2 + M$  ruter fra lag 0 til og med lag  $(N-1)$ . Vi legger sammen 2 strekninger i hver rute, pluss at vi i siste lag  $N$  har bare 1 valg per by og dermed  $M$  ruter. Totalt antall ruter blir da  $2(N-1)M^2 + 3M$  som er mye mindre enn  $NM^N$  som vi fant først.

## 2.2 Hamilton-Jacobi-ligningen

Vi fortsetter med tilsvarende problem, med fastsatt endetid  $T$ , som vi løste med Pontryagins maksimalprinsipp. Vi har et dynamiske system:

$$\begin{cases} \dot{x}(t) = f(x(t), \alpha(t)) & t \in [0, \infty) \\ x(0) = x_0 & x_0 \in \mathbb{R}^d \end{cases} \quad (35)$$

med en tilhørende avkastningsfunksjonal:

$$P[\alpha(\cdot)] := \int_0^T r(x(t), \alpha(t)) dt + g(x(T))$$

Som vi ønsker å maksimere. I stedet for å løse problemet fra 0 til  $T$ , skal vi bruke ideen fra dynamisk programmering og utvide problemet til å omhandle alle  $t \in [0, T]$  og  $x \in \mathbb{R}^d$ .

Vi definerer som tidligere  $\mathcal{A} = \{\alpha(\cdot) : [0, T] \rightarrow A \mid \alpha(\cdot) \text{ er målbar}\}$  som mengden med alle mulige kontroller.

**Definisjon 2.2.** For  $x \in \mathbb{R}^d$  og  $t \in [0, T]$  definerer vi verdifunksjonen  $v(t, x)$  som den største mulige avkastningen når vi starter i  $x \in \mathbb{R}^n$  ved tiden  $t$ .

$$v(t, x) = \sup_{\alpha \in \mathcal{A}} P_{x,t}[\alpha(\cdot)] := \sup_{\alpha \in \mathcal{A}} \int_t^T r(x(s), \alpha(s)) dt + g(x(T)) \quad s \in [t, T]$$

Ser vi på tilfellet  $v(0, x_0)$  er dette det opprinnelige problemet, men nå er det bare et av uendelig mange delproblemer.

**Teorem 2.1.** Anta at verdifunksjonen  $v(t, x)$  er  $C^1$  i både  $t$  og  $x$ . Da løser  $v(t, x)$  den ikkelineære partielle differensialligningen vi kaller Hamilton-Jacobi-ligningen:

$$HJ \quad v_t(t, x) + \max_{a \in A} [f(x, a)v_x(t, x) + r(x, a)] = 0 \quad (x \in \mathbb{R}^n, t \in [0, T]) \quad (36)$$

Med betingelsen:

$$v(T, x) = g(x)$$

*Bevis.* La nå  $x \in \mathbb{R}^n$ ,  $h > 0$  og  $0 \leq t \leq T$  være gitt.  $\mathcal{A}$  er mengden mulige kontroller  $\alpha(\cdot)$ .

Vi skal nå utnytte at  $v(t, x)$  er den optimale løsningen på hele intervallet  $[t, T]$ . Først velger vi en konstant kontroll  $\alpha(\cdot) \equiv a \in A$  på intervallet  $t \leq s \leq t+h$  slik at  $t+h < T$ . På det resterende intervallet  $[t+h, T]$  velger vi den optimale kontrollen  $\alpha^*(\cdot)$ .

Nå er det dynamiske systemet:

$$\begin{cases} \dot{x}(s) = f(x(s), a) & s \in [t, t+h] \\ \dot{x}(s) = f(x(s), \alpha^*(s)) & s \in [t+h, T] \\ x(t) = x \end{cases} \quad (37)$$

Avkastningsfunksjonale blir nå:

$$\int_t^{t+h} r(x(s), a) ds + \int_{t+h}^T r(x(s), \alpha^*(s)) ds + g(x(T)) = \int_t^{t+h} r(x(s), a) ds + v(x(t+h), t+h)$$

På hele intervallet  $[t, T]$  er  $v(t, x)$  definert som supremum over alle  $\alpha(\cdot) \in \mathcal{A}$  så av dette følger det at:

$$v(t, x) \geq \int_t^{t+h} r(x(s), a) ds + v(t+h, x(t+h))$$

Stokker vi om på uttrykket får vi:

$$v(t+h, x(t+h)) - v(t, x) + \int_t^{t+h} r(x(s), a) ds \leq 0$$

Vi kan så dele på  $h$  og få:

$$\frac{v(t+h, x(t+h)) - v(t, x)}{h} + \frac{1}{h} \int_t^{t+h} r(x(s), a) ds \leq 0$$

La  $h \rightarrow 0$  så får vi i første ledd:

$$\lim_{h \rightarrow 0} \frac{v(t+h, x(t+h)) - v(t, x)}{h} = \partial_t v(t, x) + \nabla_x v(t, x) \dot{x}(t) = \partial_t v(t, x) + \nabla_x v(t, x) f(x, a)$$

Siste likheten kommer av at vi bytter  $\dot{x}(t)$  med  $f(x(t), a)$  fra (37). Da har vi:

$$\partial_t v(t, x) + \nabla_x v(t, x) f(x, a) + r(x, a) \leq 0$$

Denne ulikheten holder for alle  $a \in A$  så vi kan skrive:

$$\partial_t v(t, x) + \max_{a \in A} \{ \nabla_x v(t, x) f(x, a) + r(x, a) \} \leq 0$$

Anta nå at vi har valgt  $\alpha^*(\cdot)$  slik at vi finner maksimum. Det følger da en tilhørende løsning  $x^*(s)$  av det dynamiske systemet(37) og:

$$v(t, x) = \int_t^{t+h} r(x^*(s), \alpha^*(s)) ds + v(t+h, x^*(t+h))$$

Etter tilsvarende opprydning og at vi deler på  $h$  som over får vi:

$$\frac{v(t+h, x^*(t+h)) - v(t, x)}{h} + \frac{1}{h} \int_t^{t+h} r(x^*(s), \alpha^*(s)) ds = 0$$

Lar så  $h \rightarrow 0$  og antar at  $\alpha^*(t) = a^*$  Da er:

$$\partial_t v(t, x) + \nabla_x v(t, x) f(x, a^*) + r(x, a^*) = 0$$

Som viser at  $v(t, x)$  løser Hamilton-Jacobi-ligningen. □

### 2.2.1 Konstruksjon av den optimale kontrollen $\alpha^*(\cdot)$

Vi skal nå finne den optimale kontrollen  $\alpha(\cdot)$  slik at vi for alle  $x \in \mathbb{R}^d$  og alle  $t \in [0, T]$  maksimerer  $P_{t,x}[\alpha(\cdot)]$ . Vi starter med å løse Hamilton-Jacobi-ligningen og finner  $v(t, x)$ . Nå kan vi bruke Hamilton-Jacobi-ligningen og verdifunksjonen  $v(t, x)$  til å konstruerer  $\alpha(\cdot)$ .

For alle  $x \in \mathbb{R}^d$  og alle  $t \in [0, T]$  konstruerer vi  $\alpha(x, t) = a \in A$  slik at:

$$\partial_t v(t, x) + \nabla_x v(t, x) f(x, a) + r(x, a) = 0 \tag{38}$$

Med dette valg av  $\alpha(x, t)$  kan vi, om det lar seg gjøre, løse:

$$\begin{cases} \dot{x}^*(s) = f(x^*(s), \alpha(x^*(s), s)) & t \leq s \leq T \\ x(t) = x \end{cases}$$

med denne løsningen kan vi nå definere optimale kontrollen ved:

$$\alpha^*(s) := \alpha(x^*(s), s) \tag{39}$$

### 2.2.2 Bevis for at $\alpha^*(\cdot)$ er optimal

Vi har definert:

$$P_{x,t}[\alpha^*(\cdot)] = \int_t^T r(x^*(s), \alpha^*(s)) ds + g(x^*(T))$$

Dessuten har vi fra definisjonen (38) at:

$$r(x^*(s), \alpha^*(s)) = -\partial_s v(s, x^*(s)) - \nabla_x v(s, x^*(s)) f(x^*(s), \alpha(s, x^*(s)))$$

Som vi kan sette inn i integralet:

$$\begin{aligned} P_{t,x}[\alpha^*(\cdot)] &= \int_t^T [-\partial_s v(s, x^*(s)) - \nabla_x v(s, x^*(s)) f(x^*(s), \alpha(s, x^*(s)))] ds + g(x^*(T)) \\ &= - \int_t^T [\partial_s v(s, x^*(s)) + \nabla_x v(s, x^*(s)) \dot{x}^*(s)] ds + g(x^*(T)) \\ &= - \int_t^T \frac{d}{ds} v(s, x^*(s)) ds + g(x^*(T)) \\ &= -v(T, x^*(T)) + v(t, x^*(t)) + g(x^*(T)) \\ &= v(t, x^*(t)) = \sup_{\alpha(\cdot) \in \mathcal{A}} P_{t,x}[\alpha(\cdot)] \end{aligned}$$

Så av dette følger det at:

$$P_{x,t}[\alpha^*(\cdot)] = \sup_{\alpha(\cdot) \in \mathcal{A}} P_{t,x}[\alpha(\cdot)]$$

Som var det vi ønsket å vise. □

### 3 Gradient nedstigning

Før vi viser konvergens av den stokastiske gradientalgoritmen skal vi første se hva som skjer med gradient nedstigingsalgoritmen når  $\eta \rightarrow 0$ . Fremstillingen er hentet fra (Bach,2020). Vi ønsker å minimere en funksjon  $f : \mathbb{R}^d \rightarrow \mathbb{R}$  på formen:

$$\min_{x \in \mathbb{R}^d} f(x) = \min_{x \in \mathbb{R}^d} \frac{1}{n} \sum_{i=1}^n f_i(x)$$

Hvor  $f_i : \mathbb{R}^d \rightarrow \mathbb{R}$  Dette kan være en kostfunksjon i en maskinlæringsalgoritmen, hvor  $n$  er antall data/målepunkter vi har, og  $d$  er antall parameter vi minimere over.

Når det ikke eksisterer en analytisk løsning på minimeringsproblemet kan vi bruke gradienten til funksjonen for å, steg for steg nærme oss løsningen. Det er dette som gir oss gradient nedstigingskjema med startpunkt  $x_0$ :

$$x_{k+1} = x_k - \eta \nabla f(x_k) \tag{40}$$

Lengden på stegene vi tar er  $\eta$ , også kalt læringsraten. Vi skal nå vise at når læringsraten  $\eta$  går mot null konvergerer gradientalgoritmen mot en ordinær differensialligning.

**Lemma 3.1.** La  $\eta > 0$ ,  $k \in \mathbb{N}$  og  $x_{k+1}$  er steg  $k+1$  i gradient nedstigingsalgoritmen:

$$x_{k+1} = x_k - \eta \nabla f(x_k) \quad x_k \in \mathbb{R}^d \quad \forall k$$

da konvergerer gradientalgoritmen, når  $\eta \rightarrow 0$ , mot den ordinære differensialligningen:

$$\dot{x}(t) = -\nabla f(x(t)) \quad x \in \mathbb{R}^d$$

*Bevis.* Vi skriver om  $x_k = x_{\eta k}$  og velger en kontinuerlig funksjon  $x(t)$  slik at  $x(\eta k) = x_{\eta k}$  for alle  $k$ . Velger vi nå  $t = \eta k$  følger det at:

$$x_{\eta(k+1)} = x(\eta(k+1)) = x(\eta k + \eta) = x(t + \eta)$$

Bytter vi det inn i skjema for gradientalgoritmen (40) får vi:

$$x(t + \eta) = x(t) - \eta \nabla f(x(t))$$

og:

$$\frac{x(t + \eta) - x(t)}{\eta} = -\nabla f(x(t))$$

Lar vi nå  $\eta \rightarrow 0$  får vi:

$$\lim_{\eta \rightarrow 0} \frac{x(t + \eta) - x(t)}{\eta} = \dot{x}(t) = -\nabla f(x(t))$$

Som viser at gradient nedstignings algoritmen konvergerer mot en  $d$ -dimensjonal ordinær differensialligning når læringsraten  $\eta$  går mot 0. □

### 3.1 Konvergens av stokastisk gradientalgoritmen

Vi har fremdeles det samme minimeringsproblemet som over, men erstatter gradient nedstigning med en stokastiske gradient nedstigningsalgoritme og spørsmålet er om denne algoritmen og konvergerer når  $\eta$  går mot null og hva den eventuelt konvergerer mot.

La  $D = \{\text{familie delmengder med størrelse } r \text{ fra } \{1, 2, \dots, n-1, n\}\}$  og  $\{\delta_k\}_{k \in \mathbb{N}}$  er en følge uniformt fordelt tilfeldig variabler. Vi erstatter nå gradienten i (40) med kun en tilfeldig del av gradienten og får den stokastiske gradientalgoritmen:

$$X_{k+1} = X_k - \frac{\eta}{r} \sum_{i=1}^n f_i(X_k) 1_{\delta(k+1)}(i)$$

For den stokastiske gradientalgoritmen krever det litt mer arbeid for å komme frem til et resultat. Vi skal først gjennom en omskrivning av algoritmen vise et par egenskaper ved algoritmen. Videre bruker vi et teorem hentet fra (Stroock, & Varadhan, 2006, Teorem 5.2.3 ) som viser at en Markov-kjede



konvergerer mot en diffusjon under gitte antagelser. Til slutt viser vi at den stokastiske gradientalgoritmen konvergerer i fordeling mot en løsningen av en SDE. Denne fremstillingen er en generalisering av beviset til Lanconelli og Lauria (Lanconelli, & Lauria, 2022).

Det Lanconelli og Lauria viser er at når  $r = 1$ , og under riktige betingelser på  $\nabla f(x)$  og  $\Sigma(x)$  vil:

$$X_{k+1} = X_k - \nabla f_{\delta(k+1)}(X_k)$$

konvergere i fordeling mot den svake entydige løsningen av den stokastiske differensialligning:

$$dX_t = -\nabla f(X_s)ds + \sqrt{\bar{\eta}}\Sigma(X_s)^{\frac{1}{2}}dB_s,$$

hvor

$$\Sigma(x) = \frac{1}{n} \sum_{i=1}^n (\nabla f(x) - \nabla f_i(x))(\nabla f(x) - \nabla f_i(x))^T$$

Her er  $\bar{\eta}$  en fiksering av den opprinnelige læringsraten  $\eta$  arvet fra den stokastiske gradientalgoritmen.

I beviset til Lanconelli og Lauria følges kun en av gradientene i hver iterasjon. I denne generalisert versjonen følger vi  $0 < r < n$  gradienter per iterasjon. I treningen av en maskinlæringsalgoritme vil det kunne bli arbeidskrevende å bare bruke 1

Dette kan tenkes på som størrelsen på minibatchen, man velger i mange maskinlæringsalgoritmer.

### 3.1.1 Egenskaper ved den stokastiske gradient nedstigningen

Vi har en algoritme som for hver iterasjon velger  $r$  tilfeldige datapunkter og bruker gradienten tilhørende disse for å regne ut neste iterasjon. Vi har minimeringsproblemet:

$$\min_{x \in \mathbb{R}^d} f(x) = \min_{x \in \mathbb{R}^d} \frac{1}{n} \sum_{i=1}^n f_i(x)$$

Vi velger nå en diskret stokastisk gradientalgoritme:

$$X_{k+1} = X_k - \frac{\eta}{r} \sum_{i=1}^n \nabla f_i(X_k) 1_{\delta(k+1)}(i), \quad (41)$$

hvor  $\eta > 0$  og  $\delta(k+1)$  som definert over. Det er  $\binom{n}{r}$  elementer i D. La  $\delta^j$  være et element i D. Da har vi:

$$1_{\delta^j}(i) = \begin{cases} 1 & \text{hvis } i \in \delta^j \\ 0 & \text{ellers} \end{cases} \quad i \in (1, 2, \dots, n)$$

$\delta_k$  er uniformt fordelt så  $P(\delta_k = \delta^j) = \frac{1}{\binom{n}{r}}$  for alle  $j \in (1, 2, \dots, \binom{n}{r})$ .

Vi skriver om (41) ved å legge til og trekke fra  $\eta \nabla f(X_k)$ :

$$X_{k+1} = X_k - \eta \nabla f(X_k) + \eta \nabla f(X_k) - \frac{\eta}{r} \sum_{i=1}^n \nabla f_i(X_k) 1_{\delta(k+1)}(i)$$

Høyre side kan vi nå dele i to, en deterministisk del og en del som inneholder noe tilfeldig:

$$\begin{aligned} X_{k+1} &= X_k - \eta \nabla f(X_k) + \eta (\nabla f(X_k) - \frac{1}{r} \sum_{i=1}^n \nabla f_i(X_k) 1_{\delta(k+1)}(i)) \\ &= X_k - \eta \nabla f(X_k) + \sqrt{\eta} V_k \end{aligned}$$

Vi har nå valgt  $V_k$  som en  $d$ -dimensjonal tilfeldig vektor:

$$V_k(X_k) := \sqrt{\eta} \left( \nabla f(X_k) - \frac{1}{r} \sum_{i=1}^n \nabla f_i(X_k) 1_{\delta(k+1)}(i) \right)$$

**Proposisjon 3.1.** *La  $\mathcal{F}_k$  være  $\sigma$ -algebraen generert av de tilfeldige vektorene  $X_0, X_1, \dots, X_k$ . Da er:*

$$\mathbb{E}[V_k | \mathcal{F}_k] = 0$$

*Bevis.*

$$\begin{aligned} \mathbb{E}[V_k | \mathcal{F}_k] &= \mathbb{E} \left[ \sqrt{\eta} \left( \nabla f(X_k) - \frac{1}{r} \sum_{i=1}^n \nabla f_i(X_k) 1_{\delta(k+1)}(i) \right) \middle| \mathcal{F}_k \right] \\ &= \sqrt{\eta} \left( \mathbb{E} \left[ \nabla f(X_k) \middle| \mathcal{F}_k \right] - \mathbb{E} \left[ \frac{1}{r} \sum_{i=1}^n \nabla f_i(X_k) 1_{\delta(k+1)}(i) \middle| \mathcal{F}_k \right] \right) \\ &= \sqrt{\eta} \left( \nabla f(X_k) - \mathbb{E} \left[ \frac{1}{r} \sum_{i=1}^n \nabla f_i(X_k) 1_{\delta(k+1)}(i) \right] \right) \end{aligned} \quad (42)$$

Bruker at  $\nabla f(X_k)$  er målbar m.h.p.  $\mathcal{F}_k$  og at  $\delta_{(k+1)}$  er uavhengig av  $\mathcal{F}_k$ . Videre har vi:

$$\begin{aligned} \mathbb{E} \left[ \frac{1}{r} \sum_{i=1}^n \nabla f_i(X_k) 1_{\delta(k+1)}(i) \right] &= \sum_{j=1}^{\binom{n}{r}} \frac{1}{r} \sum_{i=1}^n \nabla f_i(X_k) 1_{\delta_{(k+1)}^j}(i) \frac{1}{\binom{n}{r}} \\ &= \frac{1}{r \binom{n}{r}} \sum_{j=1}^{\binom{n}{r}} \sum_{i=1}^n \nabla f_i(X_k) 1_{\delta_{(k+1)}^j}(i) \end{aligned} \quad (43)$$

Fikser  $i \in (1, 2, \dots, n)$ . Når vi plukker ut alle  $\binom{n}{r}$  mulige kombinasjoner av vektorer med  $\nabla f_i$  medfører det at  $\nabla f_i$  finnes i  $\binom{n-1}{r-1}$  antall av disse kombinasjonene. Vi kan skrive dette om:

$$\binom{n-1}{r-1} = \frac{(n-1)!}{(r-1)!(n-1-(r-1))!} = \frac{(n-1)!}{(r-1)!(n-r)!} = \frac{r(n-1)!n}{r(r-1)!(n-r)!n} = \frac{r \binom{n}{r}}{n}$$

Vi skriver om dobbeltsummen:

$$\begin{aligned} \sum_{j=1}^{\binom{n}{r}} \sum_{i=1}^n \nabla f_i(X_k) 1_{\delta_j}(i) &= (\nabla f_1 + \dots + \nabla f_1) + \dots + (\nabla f_n + \dots + \nabla f_n) \\ &= \frac{r \binom{n}{r}}{n} \nabla f_1 + \dots + \frac{r \binom{n}{r}}{n} \nabla f_n = \frac{r \binom{n}{r}}{n} \sum_{i=1}^n \nabla f_i = r \binom{n}{r} \frac{1}{n} \sum_{i=1}^n \nabla f_i = r \binom{n}{r} \nabla f \end{aligned}$$

Setter vi dette tilbake i (43) og setter det tilbake i (42) får vi:

$$\sqrt{\eta} \left( \nabla f(X_k) - \mathbb{E} \left[ \frac{1}{r} \sum_{i=1}^n \nabla f_i(X_k) 1_{\delta_{(k+1)}}(i) \right] \right) = \sqrt{\eta} \left( \nabla f(X_k) - \frac{1}{r \binom{n}{r}} r \binom{n}{r} \nabla f(X_k) \right) = 0$$

□

Videre får vi fra uavhengighet og fra definisjonen den betingede  $d \times d$  kovariansmatrisen:

$$\mathbb{E}[V_k(V_k)^T | \mathcal{F}_k] = \frac{\eta}{\binom{n}{k}} \sum_{j=1}^{\binom{n}{k}} \left[ \left( \nabla f(X_k) - \frac{1}{r} \sum_{i=1}^n \nabla f_i(X_k) 1_{\delta_j}(i) \right) \left( \nabla f(X_k) - \frac{1}{r} \sum_{i=1}^n \nabla f_i(X_k) 1_{\delta_j}(i) \right)^T \right]$$

Vi definerer nå:

$$\Sigma(X) := \frac{1}{\binom{n}{k}} \sum_{j=1}^{\binom{n}{k}} \left( \nabla f(X) - \frac{1}{r} \sum_{i=1}^n \nabla f_i(X) 1_{\delta_j}(i) \right) \left( \nabla f(X) - \frac{1}{r} \sum_{i=1}^n \nabla f_i(X) 1_{\delta_j}(i) \right)^T \quad (44)$$

Da er  $\mathbb{E}[V_k(V_k)^T | \mathcal{F}_k] = \eta \Sigma(X_k)$ .

Vi velger nå å fiksure  $\eta = \bar{\eta}$  og bytter vi  $V_k$  med  $\bar{V}_k$ . Dette er en måte å huske hvor vi kommer fra, noe vi får bruk for senere når vi skal konstruere en alternativ algoritme. Vi har nå  $\bar{V}_k$ :

$$\bar{V}_k := \sqrt{\bar{\eta}} \left( \nabla f(X_k) - \frac{1}{r} \sum_{i=1}^n \nabla f_i(X_k) 1_{\delta_{(k+1)}}(i) \right)$$

Den betingede forventningen til  $\bar{V}_k$  gitt  $\mathcal{F}_k$  er:

$$\mathbb{E}[\bar{V}_k | \mathcal{F}_k] = 0 \quad (45)$$

og vi får tilhørende betingnede kovarians matrisen:

$$\mathbb{E}[\overline{V}_k(\overline{V}_k)^T | \mathcal{F}_k] = \overline{\eta} \Sigma(X_k) \quad (46)$$

Vi setter  $\overline{V}_k(X_k)$  tilbake i (41) og får.

$$X_{k+1} = X_k - \eta \nabla f(X_k) + \sqrt{\overline{\eta}} \overline{V}_k(X_k) \quad (47)$$

Med dette utgangspunktet skal vi vise at det stokastiske gradientalgoritmen konvergerer mot en stokastisk prosess.

**Teorem 3.2.** *Anta at:*

1. Gradientene  $\nabla f_i$   $i \in (1, 2, \dots, n)$  er Lipschitz kontinuerlige;
2. Kovariansmatrisen  $\Sigma$  som definert i (44) tilfredstiller begresningen:

$$\max_{j \in \{1, \dots, d\}} \left| \left\langle \theta, \partial_{x_j}^2 \Sigma(x) \theta \right\rangle \right| \leq \lambda_0 \|\theta\|^2, \quad \theta \in \mathbb{R}^d \quad (48)$$

for alle  $x \in \mathbb{R}^d$ , Operatoren  $\partial_{x_j}^2$  gjelder elementvis i matrisen.

Under disse betingelsen konvergerer den diskret stokastiske prosessen  $\{X_k\}_{k \geq 0}$ ,  $k \in \mathbb{N}$  (41) når  $\eta \rightarrow 0$  i fordeling mot den unike svake løsningen av den stokastiske differensialligningen:

$$dX_t = -\nabla f(X_t) dt + (\overline{\eta} \Sigma(X_t))^{\frac{1}{2}} dB_t \quad (49)$$

$B_t$  er en  $d$ -dimensjonal Brownsk bevegelse.

For å bevise teorem 3.2 trenger vi først en versjon av nevnte teorem fra Stroock og Varadhan (stroock & Varadhan, 2006, Teorem 5.2.3). I likhet med Lanconelli og Lauria (Lanconelli & Lauria, 2022) benytter vi oss av en forenklet versjon. Teoremet viser at en Markov-kjede under de riktige betingelsene konvergerer mot en diffusjon.

### 3.1.2 Et generelt funksjonalgrenseteorem

La først  $\{x_{\eta k}^{(\eta)}\}_{k \geq 0}$  være en  $d$ -dimensjonal diskret, i tid, Markov-prosess på et komplett sannsynlighetsrom  $(\Omega, \mathcal{F}, P)$  indeksert med  $\eta \in \mathbb{R}^+$ . Vi antar og videre at:

$$P(x_0^{(\eta)} = x_0) = 1 \quad \forall \eta > 0 \quad (50)$$

Hvor  $x_0$  er en  $d$ -dimensjonal deterministisk vektor. Vi definerer for alle  $k > 0$  og  $\eta > 0$  en  $\sigma$ -algebra  $\mathcal{F}_{k\eta}^{(\eta)}$  generert av de tilfeldige vektorene:  $x_0, x_{\eta}^{(\eta)}, x_{2\eta}^{(\eta)}, \dots, x_{k\eta}^{(\eta)}$ . Hvor Markov-egenskapen 1.14 holder.

La  $\{X_t^{(\eta)}\}_{t \geq 0}$  være en kontinuerlig prosess i tid konstruert slik at for den diskret prosessen  $\{x_{\eta k}^{(\eta)}\}_{k \geq 0}$  har vi:

$$X_t^{(\eta)} := x_{k\eta}^{(\eta)}, \quad \text{hvis } t \in [k\eta, (k+1)\eta) \quad (51)$$

Dersom vi oppfyller de 3 påfølgende antagelsene vil dette medføre at prosessen  $\{X_t^{(\eta)}\}_{t \geq 0}$  konvergerer mot en svak løsning av en SDE når  $\eta$  går mot 0.

**Antagelse 1:** Det eksisterer kontinuerlige funksjoner  $a : \mathbb{R}^d \times [0, \infty) \rightarrow S_d$ , hvor  $S_d$  er rommet av  $d \times d$  symmetriske ikke-negativt definitte matriser, og  $b : \mathbb{R}^d \times [0, \infty) \rightarrow \mathbb{R}^d$  slik at for alle  $R > 0$  og  $T > 0$  så får vi:

$$\lim_{\eta \rightarrow 0} \sup_{\|x\| \leq R, 0 \leq t \leq T} \left\| \frac{\mathbb{E}[X_{t+\eta}^{(\eta)} - X_t^{(\eta)} | X_t^{(\eta)} = x]}{\eta} - b(x, t) \right\| = 0 \quad (52)$$

og:

$$\lim_{\eta \rightarrow 0} \sup_{\|x\| \leq R, 0 \leq t \leq T} \left\| \frac{\mathbb{E}[(X_{t+\eta}^{(\eta)} - X_t^{(\eta)})(X_{t+\eta}^{(\eta)} - X_t^{(\eta)})^T | X_t^{(\eta)} = x]}{\eta} - a(x, t) \right\| = 0 \quad (53)$$

Dessuten eksisterer det en positiv  $\beta$  slik at for alle  $j = 1, \dots, d$  har vi:

$$\lim_{\eta \rightarrow 0} \sup_{\|x\| \leq R, 0 \leq t \leq T} \frac{\mathbb{E}[|\langle X_{t+\eta}^{(\eta)} - X_t^{(\eta)}, e_j \rangle|^{2+\beta} | X_t^{(\eta)} = x]}{\eta} = 0 \quad (54)$$

Hvor  $e = \{e_1, e_2, \dots, e_j\}$  er basis i  $\mathbb{R}^d$ .

**Antagelse 2:** Det eksisterer en kontinuerlig avbildning  $\sigma : \mathbb{R}^d \times [0, \infty) \rightarrow S_d$  slik at for alle  $x \in \mathbb{R}^d$  og  $t \geq 0$  så har vi  $a(x, t) = \sigma(x, t)\sigma(x, t)^T$

**Antagelse 3:** Svak entydighet holder for SDE-en:

$$\begin{cases} dX_t &= b(X_t, t)dt + \sigma(X_t, t)dB_t & t > 0; \\ X_0 &= x_0 \end{cases} \quad (55)$$

**Teorem 3.3.** Under antagelsene (1-3) har vi at den kontinuerlige prosessen  $\{X_t^{(\eta)}\}_{t \geq 0}$ , definert som (51) konvergerer i fordeling når  $\eta$  går mot null mot den unike svake løsningen  $X_t$  av (55).

### 3.2 Bevis teorem 3.2 (for en generell stokastisk gradientalgoritme)

*Bevis.* Vi bruker funksjonalgrenseteoremet 3.3 og sjekker at antagelsene over holder for den stokastiske gradientalgoritmen når læringsraten  $\eta$  går mot 0. Vi starter med antagelse 1 og bruker  $X_t^{(\eta)}$ , fra (51) og  $x_{\eta(k+1)}^{(\eta)}$  fra (47). La  $t > 0$  være fiksert. Da finnes en  $k \in \mathbb{N}$  slik at  $k\eta \leq t < (k+1)\eta$  og vi får:

$$\begin{aligned} X_{t+\eta}^{(\eta)} - X_t^{(\eta)} &= x_{(k+1)\eta}^{(\eta)} - x_{k\eta}^{(\eta)} \\ &= -\nabla f(x_{k\eta}^{(\eta)})\eta + \overline{V_{k\eta}^{(\eta)}}\sqrt{\eta} \end{aligned} \quad (56)$$

Derfor får vi fra første antagelse:

$$\begin{aligned} \frac{\mathbb{E}[X_{t+\eta}^{(\eta)} - X_t^{(\eta)} | X_t^{(\eta)} = x]}{\eta} &= \frac{\mathbb{E}[-\nabla f(x_{k\eta}^{(\eta)})\eta + \overline{V_{k\eta}^{(\eta)}}\sqrt{\eta} | x_{k\eta}^{(\eta)} = x]}{\eta} \\ &= \mathbb{E}[-\nabla f(x_{k\eta}^{(\eta)}) | x_{k\eta}^{(\eta)} = x] + \frac{\mathbb{E}[\overline{V_{k\eta}^{(\eta)}} | x_{k\eta}^{(\eta)} = x]}{\sqrt{\eta}} \end{aligned} \quad (57)$$

Vi bruker at  $\nabla f(x_{k\eta}^{(\eta)})$  er målbar og fra (45) har vi  $\mathbb{E}[\overline{V_{k\eta}^{(\eta)}} | x_{k\eta}^{(\eta)} = x] = 0$ , dermed har vi:

$$\frac{\mathbb{E}[X_{t+\eta}^{(\eta)} - X_t^{(\eta)} | X_t^{(\eta)} = x]}{\eta} = -\nabla f(x) \quad (58)$$

Om vi velger  $b(x, t) = -\nabla f(x)$  har vi oppfylt (52).

For å sjekke (53) bruker vi tilsvarende omskrivning som over:

$$\begin{aligned} &\frac{\mathbb{E}[(X_{t+\eta}^{(\eta)} - X_t^{(\eta)})(X_{t+\eta}^{(\eta)} - X_t^{(\eta)})^T | X_t^{(\eta)} = x]}{\eta} = \\ &\frac{\mathbb{E}[(-\nabla f(x_{k\eta}^{(\eta)})\eta + \overline{V_{k\eta}^{(\eta)}}\sqrt{\eta})(-\nabla f(x_{k\eta}^{(\eta)})\eta + \overline{V_{k\eta}^{(\eta)}}\sqrt{\eta})^T | x_{k\eta}^{(\eta)} = x]}{\eta} \\ &= \frac{1}{\eta} \left( \mathbb{E} \left[ \eta^2 \nabla f(x_{k\eta}^{(\eta)}) (\nabla f(x_{k\eta}^{(\eta)}))^T - \nabla f(x_{k\eta}^{(\eta)}) \eta (\overline{V_{k\eta}^{(\eta)}}\sqrt{\eta})^T - \right. \right. \\ &\quad \left. \left. \overline{V_{k\eta}^{(\eta)}}\sqrt{\eta} (\nabla f(x_{k\eta}^{(\eta)})\eta)^T + \overline{V_{k\eta}^{(\eta)}}\sqrt{\eta} (\overline{V_{k\eta}^{(\eta)}}\sqrt{\eta})^T | x_{k\eta}^{(\eta)} = x \right] \right) \\ &= \mathbb{E} \left[ \eta \nabla f(x_{k\eta}^{(\eta)}) (\nabla f(x_{k\eta}^{(\eta)}))^T | x_{k\eta}^{(\eta)} = x \right] - \sqrt{\eta} \mathbb{E} \left[ \nabla f(x_{k\eta}^{(\eta)}) (\overline{V_{k\eta}^{(\eta)}})^T | x_{k\eta}^{(\eta)} = x \right] - \\ &\quad \sqrt{\eta} \mathbb{E} \left[ \overline{V_{k\eta}^{(\eta)}} (\nabla f(x_{k\eta}^{(\eta)}))^T | x_{k\eta}^{(\eta)} = x \right] + \mathbb{E} \left[ \overline{V_{k\eta}^{(\eta)}} (\overline{V_{k\eta}^{(\eta)}})^T | x_{k\eta}^{(\eta)} = x \right] \\ &= \eta \nabla f(x) \nabla f(x)^T - \nabla f(x) \sqrt{\eta} \mathbb{E} \left[ (\overline{V_{k\eta}^{(\eta)}})^T | x_{k\eta}^{(\eta)} = x \right] - \\ &\quad \sqrt{\eta} \mathbb{E} \left[ \overline{V_{k\eta}^{(\eta)}} | x_{k\eta}^{(\eta)} = x \right] \nabla f(x)^T + \mathbb{E} \left[ \overline{V_{k\eta}^{(\eta)}} (\overline{V_{k\eta}^{(\eta)}})^T | x_{k\eta}^{(\eta)} = x \right] \end{aligned}$$

Andre og tredje ledd er null fra (45). Og i siste overgang har vi brukt egenskap nummer 5 fra den betingede forventningen som gjør at vi kan flytte ut  $\nabla f(x)$ . Lar vi nå  $\eta$  gå mot null får vi dermed:

$$\begin{aligned} \lim_{\eta \rightarrow 0} \frac{\mathbb{E}[(X_{t+\eta}^{(\eta)} - X_t^{(\eta)})(X_{t+\eta}^{(\eta)} - X_t^{(\eta)})^T | X_t^{(\eta)} = x]}{\eta} &= \\ \lim_{\eta \rightarrow 0} \left[ \eta \nabla f(x) \nabla f(x)^T + \mathbb{E} \left[ \overline{V_{k\eta}^{(\eta)}} (\overline{V_{k\eta}^{(\eta)}})^T | x_{k\eta}^{(\eta)} = x \right] \right] &= \bar{\eta} \Sigma(x) \end{aligned} \quad (59)$$

Vi er her avhengig av betingelsen at  $\nabla f(x)$  er antatt Lipschitz kontinuert. Da er  $\nabla f(x) \nabla f(x)^T$  begrenset på en kompakt mengde og dette leddet går mot 0 når  $\eta$  går mot 0. Velger vi  $a(x, t) = \bar{\eta} \Sigma(x)$  som definert over tilfredstiller dette

(53) i den første antagelsen.

Før vi går løst på den siste likheten i den første antagelsen trenger vi et lite miniresultat.

**Lemma 3.4.** *La  $p \in [1, \infty)$  da holder:*

$$|a + b|^p \leq 2^{p-1}(|a|^p + |b|^p)$$

*Bevis.* La  $f(x) = x^p$  for  $x > 0$ . Da er  $f$  konveks ettersom den andrederiverte  $f''(x) = p(p-1)x^{p-2} \geq 0$ . Fra dette følger ved bruk av trekantulikheten og Jensens ulikhet:

$$\left(\frac{|a+b|}{2}\right)^p \leq \left(\frac{|a|+|b|}{2}\right)^p \leq \frac{|a|^p + |b|^p}{2} \implies |a+b|^p \leq 2^{p-1}(|a|^p + |b|^p)$$

□

Velg en  $j \in (1, 2, \dots, n)$  og la  $\beta > 0$  da får vi ved å bruke at indreproduktet  $\langle \cdot, \cdot \rangle$  er lineært i første faktor:

$$\begin{aligned} \langle X_{t+\eta}^{(\eta)} - X_t^{(\eta)}, e_j \rangle &= \langle -\nabla f(x_{k\eta}^{(\eta)})\eta + \sqrt{\eta} \overline{V_{k\eta}^{(\eta)}}, e_j \rangle = -\eta \langle \nabla f(x_{k\eta}^{(\eta)}), e_j \rangle + \sqrt{\eta} \langle \overline{V_{k\eta}^{(\eta)}}, e_j \rangle \\ &= -\eta \partial_{x_j} f(x_{k\eta}^{(\eta)}) + \sqrt{\eta} \langle \sqrt{\eta} (\nabla f(x_{k\eta}^{(\eta)}) - \frac{1}{r} \sum_{i=1}^n \nabla f_i(x_{k\eta}^{(\eta)}) 1_{\delta(k+1)}(i)), e_j \rangle \\ &= -\eta \partial_{x_j} f(x_{k\eta}^{(\eta)}) + \sqrt{\eta} \sqrt{\eta} \partial_{x_j} f(x_{k\eta}^{(\eta)}) - \sqrt{\eta} \sqrt{\eta} \frac{1}{r} \sum_{i=1}^n \partial_{x_j} f_i(x_{k\eta}^{(\eta)}) 1_{\delta(k+1)}(i) \end{aligned}$$

Dette gir:

$$\begin{aligned} &\frac{\mathbb{E} \left[ |\langle X_{t+\eta}^{(\eta)} - X_t^{(\eta)}, e_j \rangle|^{2+\beta} \mid x_{k\eta}^{(\eta)} = x \right]}{\eta} = \\ &\frac{\mathbb{E} \left[ \left| -\eta \partial_{x_j} f(x_{k\eta}^{(\eta)}) + \sqrt{\eta} \sqrt{\eta} \partial_{x_j} f(x_{k\eta}^{(\eta)}) - \sqrt{\eta} \sqrt{\eta} \frac{1}{r} \sum_{i=1}^n \partial_{x_j} f_i(x_{k\eta}^{(\eta)}) 1_{\delta(k+1)}(i) \right|^{2+\beta} \mid x_{k\eta}^{(\eta)} = x \right]}{\eta} \end{aligned}$$

Bruker lemma 3.4:

$$\begin{aligned} &\leq 2^{1+\beta} \frac{\mathbb{E} \left[ \left| -\eta \partial_{x_j} f(x_{k\eta}^{(\eta)}) \right|^{2+\beta} + \left| \sqrt{\eta} \sqrt{\eta} \partial_{x_j} f(x_{k\eta}^{(\eta)}) - \sqrt{\eta} \sqrt{\eta} \frac{1}{r} \sum_{i=1}^n \partial_{x_j} f_i(x_{k\eta}^{(\eta)}) 1_{\delta(k+1)}(i) \right|^{2+\beta} \mid x_{k\eta}^{(\eta)} = x \right]}{\eta} \\ &= 2^{1+\beta} \left[ \frac{\mathbb{E} \left[ \left| -\eta \partial_{x_j} f(x_{k\eta}^{(\eta)}) \right|^{2+\beta} \mid x_{k\eta}^{(\eta)} = x \right]}{\eta} + \right. \\ &\quad \left. \frac{\mathbb{E} \left[ \left| \sqrt{\eta} \sqrt{\eta} \partial_{x_j} f(x_{k\eta}^{(\eta)}) - \sqrt{\eta} \sqrt{\eta} \frac{1}{r} \sum_{i=1}^n \partial_{x_j} f_i(x_{k\eta}^{(\eta)}) 1_{\delta(k+1)}(i) \right|^{2+\beta} \mid x_{k\eta}^{(\eta)} = x \right]}{\eta} \right] \end{aligned}$$

$$= -2^{1+\beta} \eta^{1+\beta} |\partial_{x_j} f(x)|^{1+\beta} + \frac{2^{2(1+\beta)} \mathbb{E} \left[ \left| \sqrt{\eta} \sqrt{\bar{\eta}} \partial_{x_j} f(x_{k\eta}^{(\eta)}) - \sqrt{\eta} \sqrt{\bar{\eta}} \frac{1}{r} \sum_{i=1}^n \partial_{x_j} f_i(x_{k\eta}^{(\eta)}) 1_{\delta(k+1)}(i) \right|^{2+\beta} \middle| x_{k\eta}^{(\eta)} = x \right]}{\eta}$$

Bruker lemma 3.4 igjen:

$$\begin{aligned} &\leq 2^{1+\beta} \eta^{1+\beta} |\partial_{x_j} f(x)|^{1+\beta} + \frac{2^{2(1+\beta)} \mathbb{E} \left[ \left| \sqrt{\eta} \sqrt{\bar{\eta}} \partial_{x_j} f(x_{k\eta}^{(\eta)}) \right|^{2+\beta} + \left| \sqrt{\eta} \sqrt{\bar{\eta}} \frac{1}{r} \sum_{i=1}^n \partial_{x_j} f_i(x_{k\eta}^{(\eta)}) 1_{\delta(k+1)}(i) \right|^{2+\beta} \middle| x_{k\eta}^{(\eta)} = x \right]}{\eta} \\ &= 2^{1+\beta} \eta^{1+\beta} |\partial_{x_j} f(x)|^{1+\beta} + 2^{2(1+\beta)} \eta^{\frac{\beta}{2}} \bar{\eta}^{1+\frac{\beta}{2}} \left( \mathbb{E} \left[ \left| \partial_{x_j} f(x_{k\eta}^{(\eta)}) \right|^{2+\beta} \middle| x_{k\eta}^{(\eta)} = x \right] + \mathbb{E} \left[ \left| \frac{1}{r} \sum_{i=1}^n \partial_{x_j} f_i(x_{k\eta}^{(\eta)}) 1_{\delta(k+1)}(i) \right|^{2+\beta} \middle| x_{k\eta}^{(\eta)} = x \right] \right) \\ &= 2^{1+\beta} \eta^{1+\beta} |\partial_{x_j} f(x)|^{1+\beta} + 2^{2(1+\beta)} \eta^{\frac{\beta}{2}} \bar{\eta}^{1+\frac{\beta}{2}} \left( |\partial_{x_j} f(x)|^{2+\beta} + \mathbb{E} \left[ \left| \frac{1}{r} \sum_{i=1}^n \partial_{x_j} f_i(x_{k\eta}^{(\eta)}) 1_{\delta(k+1)}(i) \right|^{2+\beta} \middle| x_{k\eta}^{(\eta)} = x \right] \right) \\ &\leq 2^{1+\beta} \eta^{1+\beta} |\partial_{x_j} f(x)|^{1+\beta} + 2^{2(1+\beta)} \eta^{\frac{\beta}{2}} \bar{\eta}^{1+\frac{\beta}{2}} \left( |\partial_{x_j} f(x)|^{2+\beta} + \frac{1}{r} \sum_{i=1}^n \mathbb{E} \left[ \left| \partial_{x_j} f_i(x_{k\eta}^{(\eta)}) 1_{\delta(k+1)}(i) \right|^{2+\beta} \middle| x_{k\eta}^{(\eta)} = x \right] \right) \\ &= 2^{1+\beta} \eta^{1+\beta} |\partial_{x_j} f(x)|^{1+\beta} + 2^{2(1+\beta)} \eta^{\frac{\beta}{2}} \bar{\eta}^{1+\frac{\beta}{2}} \left( |\partial_{x_j} f(x)|^{2+\beta} + \frac{1}{r} \sum_{i=1}^n \mathbb{E} \left[ \left| \partial_{x_j} f_i(x) 1_{\delta(k+1)}(i) \right|^{2+\beta} \right] \right) \\ &\leq 2^{1+\beta} \eta^{1+\beta} |\partial_{x_j} f(x)|^{1+\beta} + 2^{2(1+\beta)} \eta^{\frac{\beta}{2}} \bar{\eta}^{1+\frac{\beta}{2}} \left( |\partial_{x_j} f(x)|^{2+\beta} + \frac{1}{r \binom{n}{r}} \sum_{i=1}^n \sum_{j=1}^{\binom{n}{r}} \left| \partial_{x_j} f_i(x) 1_{\delta_{(k+1)}^j}(i) \right|^{2+\beta} \right) \end{aligned}$$

Dette medfører at:

$$\lim_{\eta \rightarrow 0} \frac{\mathbb{E} \left[ \left| \langle X_{t+\eta}^{(\eta)} - X_t^{(\eta)}, e_j \rangle \right|^{2+\beta} \middle| x_{k\eta}^{(\eta)} = x \right]}{\eta} = 0$$

Ettersom  $\nabla f(x)$  er antatt Lipchitz og dermed begrenset på kompakte mengder  $\|x\| \leq R$ .

For antagelse 2 har vi allerede funnet at  $a(t, x) = \bar{\eta} \Sigma(x)$ . Dessuten har vi at  $\Sigma(x)$  er kovariansmatrise som medfører at den er symmetrisk positiv semidefinit.

**Proposisjon 3.2.** *La  $\Sigma$  være en symmetrisk positiv semidefinit  $d \times d$  matrise. Da finnes en unik  $d \times d$  matrise  $\sigma$  som er symmetrisk og positiv semidefinit slik at:*

$$\sigma \sigma = \Sigma$$

*Bevis.* Vi viser eksistens først:

Siden  $\Sigma$  er symmetrisk kan vi dekomponere  $\Sigma$ . La  $D$  være diagonalmatrisen med reelle egenverdiene  $\lambda_i$  og  $U$  matrisen med alle tilhørende egenvektorer.



Da er  $\Sigma = UDU^T$ . Egenverdiene  $\lambda_i \geq 0$  for alle  $i \in (1, 2, \dots, d)$  og  $U^{-1} = U^T$  siden  $\Sigma$  er positiv semidefinit. La nå  $\sqrt{D}$  være diagonalmatrisen med  $\sqrt{\lambda_i} = \{$  den positive roten til  $\lambda_i\}$  for alle  $i$  hvor  $\lambda_i \neq 0$  og  $(\sqrt{\lambda})_i = 0$  for alle  $i$  hvor  $\lambda_i = 0$ . Da er  $\sqrt{D}\sqrt{D} = D$  Vi kan nå dekomponere  $\Sigma$ :

$$\begin{aligned}\Sigma &= UDU^T = U\sqrt{D}\sqrt{D}U^{-1} \\ &= (U\sqrt{D}U^{-1})U\sqrt{D}U^{-1} = \sigma\sigma\end{aligned}$$

Unikhet: Anta at det finnes  $\sigma_1$  og  $\sigma_2$  som begge er symmetriske, positiv semidefinitte og oppfyller:

$$\sigma_1\sigma_1 = \Sigma = \sigma_2\sigma_2$$

Da har vi

$$\sigma_1\sigma_1 = UD^1U^TUD^1U^T \text{ og } UD^2U^TUD^2U^T = \sigma_2\sigma_2 \implies D^1D^1 = D^2D^2$$

Både  $D^1D^1$  og  $D^2D^2$  er diagonale med kvadrater langs diagonalen. Siden  $D^1$  og  $D^2$  er positive må de være like ettersom det kun finnes en positiv rot.  $\square$

Velger vi nå  $\sigma(x)$  til å være  $U(x)\sqrt{D(x)}U(x)^T$  fra dekomponeringen av  $\Sigma(x)$  over er denne symmetrisk positiv semidefinit. Begrensningene gjort på  $\Sigma$  i 3.2 impliserer at  $\sigma$  er globalt Lipschitz ved bruk av teorem 5.2.3 fra Stroock og Varadhan (Stroock, & Varadhan, 2006, Teorem 5.2.3) og vi har oppfylt antagelsen nummer 2.

Antagelse 3 oppfylles ettersom begrensningen som er lagt  $\nabla f(x)$  og  $\Sigma(x)$  i teorem 3.2 medfører at både  $\nabla f(x)$  og  $\sigma(x)$  er Lipschitz og vi har ikke bare en svak men en sterk løsning av SDE-en.

$$\begin{cases} dX_t = -\nabla f(X_t)dt + (\bar{\eta}\Sigma(X_t))^{\frac{1}{2}}dB_t & t > 0 \\ X_0 = x_0 \end{cases}$$

$\square$

### 3.3 Alternativ prosess $\bar{X}_t$

Den stokastiske differensialligningen vi har funnet vil i mange tilfeller være praktisk umulig å implementere når  $n$  blir stor og enda verre når  $r$  og  $\sigma$  blir av en viss størrelse. I tillegg er roten av en matrise  $\Sigma^{\frac{1}{2}}$  heller ikke entydig og det kan være uklart hvordan datamaskinen tolker denne roten. Vi velger  $\sigma(x)$  slik at  $\sigma(x)$  er positiv semidefinit, men i utgangspunktet finnes det mange matriser  $A$  slik at  $AA = \Sigma$ . La  $k$  være antall egenverdier  $\lambda_i \neq 0$  tilhørende  $\Sigma$ . Vi kan for hver  $i$  velge både  $\pm\sqrt{\lambda_i}$ . Dette gjør at vi kan konstruere  $2^k$  forskjellige matriser  $A$  slik at  $AA = \Sigma$ .

Vi skal nå vise at det finnes en alternativ prosess  $\bar{X}_t$  vi kan bruke i stedet for  $X_t$  med samme fordeling, men hvor vi ikke er nødt til å finne roten av  $\Sigma$ .

**Lemma 3.5.** La  $x_0 \in \mathbb{R}^n$  og  $X_t$  er prosessen som løser den stokastiske differensialligningen:

$$\begin{cases} dX_t &= -\nabla f(X_t)dt + (\bar{\eta}\Sigma(X_t))^{\frac{1}{2}}dB_t & t > 0 \\ X_0 &= x_0 \end{cases} \quad (60)$$

og  $\bar{X}_t$  er prosessen som løser:

$$\begin{cases} d\bar{X}_t &= -\nabla f(\bar{X}_t)dt + \sqrt{\frac{\bar{\eta}}{\binom{n}{r}}} \sum_{i=1}^{\binom{n}{r}} \left( \nabla f(\bar{X}_t) - \frac{1}{r} \sum_{j=1}^n \nabla f_j(\bar{X}_t) 1_{\delta^j}(i) \right) d\bar{B}_t^i & t > 0 \\ \bar{X}_0 &= x_0 \end{cases} \quad (61)$$

Da har  $\bar{X}_t$  og  $X_t$  samme fordeling.

For å bevise dette resultatet skal vi bruke lemma 1.6, at løsningen til Fokker-Planck-ligningen er fordelingen til den tilhørende prosessene.

*Bevis Lemma 3.5.* For at de to løsningene skal ha samme Fokker-Planck ligning må koeffisienten til driften i begge SDE-ene være lik og fra diffusjonsleddet trenger vi at  $\sigma(x)\sigma(x)^T = D$  er det samme for begge to. I både (60) og (61) er driften  $-\nabla f(x)dt$  så det holder å vise at  $D^1$  tilhørende (60) er lik  $D^2$  tilhørende (61).

For (60) har vi  $\sigma_1(x) = (\bar{\eta}\Sigma(x))^{\frac{1}{2}}$  og vi har  $\bar{\eta}\Sigma(x)^{\frac{1}{2}}(\bar{\eta}\Sigma(x)^{\frac{1}{2}})^T = \bar{\eta}\Sigma(x)$  ettersom  $\bar{\eta}\Sigma(x)^{\frac{1}{2}}$  er symmetrisk. Dette gir oss matrisen  $D^1 = \bar{\eta}\Sigma(x)$ .

I (61) kan vi se på den brownske bevegelsen som en vektor  $\bar{B}_t \in \mathbb{R}^{\binom{n}{r}}$  og diffusjonsleddet er:

$$\sqrt{\frac{\bar{\eta}}{\binom{n}{r}}} \sum_{i=1}^{\binom{n}{r}} \left( \nabla f(\bar{X}_t) - \frac{1}{r} \sum_{j=1}^n \nabla f_j(\bar{X}_t) 1_{\delta^j}(i) \right) d\bar{B}_t^i \in \mathbb{R}^d$$

. Det betyr at vi kan få det over på matrise vektorform og at det finnes en  $\sigma_2(x) \in \mathbb{R}^{d \times \binom{n}{r}}$  slik at  $\sigma_2(x)d\bar{B}_t \in \mathbb{R}^d$ . Vi får da:

$$\sqrt{\frac{\bar{\eta}}{\binom{n}{r}}} \begin{bmatrix} \partial_{x_1} f(x) - \frac{1}{r} \sum_{j=1}^n \partial_{x_1} f_j 1_{\delta^1}(i) & \dots & \partial_{x_1} f(x) - \frac{1}{r} \sum_{j=1}^n \partial_{x_1} f_j 1_{\delta^{\binom{n}{r}}}(i) \\ \partial_{x_2} f(x) - \frac{1}{r} \sum_{j=1}^n \partial_{x_2} f_j 1_{\delta^1}(i) & \dots & \vdots \\ \vdots & \ddots & \vdots \\ \partial_{x_d} f(x) - \frac{1}{r} \sum_{j=1}^n \partial_{x_d} f_j 1_{\delta^1}(i) & \dots & \partial_{x_d} f(x) - \frac{1}{r} \sum_{j=1}^n \partial_{x_d} f_j 1_{\delta^{\binom{n}{r}}}(i) \end{bmatrix} \begin{bmatrix} d\bar{B}_t^1 \\ d\bar{B}_t^2 \\ \vdots \\ d\bar{B}_t^{\binom{n}{r}} \end{bmatrix} = \sigma_2(x)d\bar{B}_t \quad (62)$$

som igjen gir oss  $D^2 = \sigma_2 \sigma_2^T$  matrisen tilhørende (61). For å vise at  $D^1 = D^2$  holder det å vise at  $D_{i,j}^1 = D_{i,j}^2$  for et vilkårlig par  $i, j \in (1, 2, \dots, d)$ . For  $D^1$  har vi fra (44):

$$D_{i,j}^1 = \bar{\eta}\Sigma_{i,j} = \frac{\bar{\eta}}{\binom{n}{r}} \sum_{k=1}^{\binom{n}{r}} \left( \partial_{x_i} f(x) - \frac{1}{r} \sum_{l=1}^n \partial_{x_i} f_l(x) 1_{\delta^k}(i) \right) \left( \partial_{x_j} f(x) - \frac{1}{r} \sum_{l=1}^n \partial_{x_j} f_l(x) 1_{\delta^k}(j) \right)$$

for  $D^2$  har vi:

$$\begin{aligned}
D_{i,j}^2 &= \frac{1}{2} \sum_{k=1}^{\binom{n}{r}} \sigma_{(2)i,k}(x) \sigma_{(2)j,k}(x) \\
&= \frac{\bar{\eta}}{\binom{n}{r}} \sum_{k=1}^{\binom{n}{r}} \left( \partial_{x_i} f(x) - \frac{1}{r} \sum_{l=1}^n \partial_{x_i} f_l(x) 1_{\delta_k}(i) \right) \left( \partial_{x_j} f(x) - \frac{1}{r} \sum_{l=1}^n \partial_{x_j} f_l(x) 1_{\delta_k}(j) \right) \\
&= \bar{\eta} \Sigma_{i,j}(x)
\end{aligned}$$

Så begge løsningen  $\bar{X}_t$  og  $X_t$  har samme Fokker-Planck ligning og har dermed sammen fordeling.  $\square$

## 4 Konstruksjon av algoritme

Vi skal nå gjennomgå steg for steg oppbygningen av den alternative stokastiske gradientalgoritmen til Li m.fl.[?] Det er og hentet fra det tilhørende supplementet til artikkelen [9]. Algoritmen baser seg på det teoretiske grunnlaget vi har gått gjennom til nå. I den stokastiske gradientalgoritmen vi bruker i dette kapitlet er antallet gradienter,  $r$  som vi følger satt til 1.  $\{\delta_k\}_{k \geq 0}$  er en følge uniformt fordelt tilfeldig variabel på mengden  $\{1, \dots, n\}$ .

Vi starter med et enkelt eksempel som illustrerer utfordringen med valg av læringsraten i den stokastisk gradientalgoritmen. La  $f(x) = x^2$  som opplagt har minimum i  $x = 0$ . Vi kan skrive  $f$  som en sum av  $f_i$ -er:

$$f(x) = \frac{1}{2} \sum_{i=1}^2 f_i(x) = \frac{1}{2} (f_1(x) + f_2(x)) = \frac{(x+1)^2 - 1}{2} + \frac{(x-1)^2 - 1}{2}$$

I tillegg har vi at:  $f'(x) = 2x$  og  $f''(x) = 2$

Vi kan med dette utgangspunktet lage et stokastisk gradientskjema med valgt  $\eta$  og  $f_i$  som definert over:

$$x_{k+1} = x_k - \eta \nabla f_{\delta(k+1)}(x)$$

Vi konstruerer  $\Sigma(x)$  på tilsvarende måte som i (44) over:

$$\Sigma(x) = \frac{1}{2} \sum_{i=1}^2 (2x - \nabla f_i(x))(2x - \nabla f_i(x)) = \frac{1}{2} ((2x - 2(x+1))^2 + (2x - 2(x-1))^2) = 4$$

$\nabla f_i(x)$  og  $\Sigma$  tilfredsstiller kriteriene i teorem 3.2 så det stokastiske gradientskjema konverger i fordeling mot løsningen av den stokastiske differensialligningen:

$$\begin{cases} dX_t &= -2X_t dt + \sqrt{\eta} dB_t \\ X_0 &= x_0 \end{cases} \quad (63)$$

Dette er en Ornstein-Uhlenbeck prosess med  $\mu = 0$ ,  $\theta = 2$  og  $\sigma = \sqrt{\eta 4}$ . Forde-  
lingen finner ved bruk av lemma 1.4:  $X_t \sim \mathcal{N}(x_0 e^{-2t}, \eta(1 - e^{-4t}))$ .

Når  $t \rightarrow \infty$  så går forventningen til  $X_t$  mot 0 som vi ønsker, men på grunn av  
støyene i variansen vil prosessen aldri slå seg helt til ro.

Vi skal nå finne en kontroll  $u(t)$  som kan styre støyen slik at når vi er nær  
minimum forsvinner støyen gradvis. Da vil prosessen stoppe opp på minimums-  
punktet.

#### 4.0.1 Innføring av kontroll i den stokastiske gradientskjema

La  $\mathcal{A} = \{u : [0, T] \rightarrow [0, 1] | u \text{ er målbar}\}$  være mengde med mulige kontroller.  
Vi innfører en kontroll,  $u_k \in \mathcal{A}$  for hver  $k$ , i det stokastiske leddet i iterasjons-  
skjema som skal kontrollere støylene. Vi får da en ny stokastisk gradientalgo-  
ritme med skjema:

$$x_{k+1} = x_k - \eta u_k \nabla f_{\delta(k+1)}(x_k) \quad (64)$$

For å komme videre gjør vi som vi som tidligere og setter  $k = \eta k$ , velger to  
kontinuerlige funksjoner i tid,  $u(t)$  og  $x(t)$  slik at  $u(\eta k) = u_{\eta k}$  og  $x(\eta k) = x_{\eta k}$   
for  $t \in [\eta k, \eta(k+1))$ . Fra teorem 3.2 konvergerer (64) i fordeling mot løsningen  
av:

$$\begin{cases} dX_t &= -u(t) \nabla f(X_t) dt + u(t) \sqrt{\eta \Sigma(x)} dB_t \\ X_0 &= x_0 \end{cases}$$

La nå  $f(x) = \frac{1}{2}a(x - b)^2$  for  $a, b \in \mathbb{R}$  og anta at  $f_i$ -ene er slik at  $\Sigma(x)$  blir kon-  
stant og større enn 0. Da vil det stokastiske gradientskjema (64) konvergere  
mot:

$$\begin{cases} dX_t &= -u(t)a(X_t - b)dt + u(t)\sqrt{\eta\Sigma}dB_t \\ X_0 &= x_0 \end{cases} \quad (65)$$

Vi antar at  $u(t) : [0, T] \rightarrow [0, 1]$  og at  $u(\cdot) \in \mathcal{A}$ .

Vi skal nå finne en optimal strategi for kontrollen  $u(\cdot)$  slik at  $X_T$  minimerer  
 $f(X_T)$ . Dette problemet er stokastisk kontrollproblem, men det kan omformu-  
leres til et deterministisk kontrollproblem hvor vi finner kontrollen  $u(\cdot)$  som  
minimerer forventningen  $\mathbb{E}[f(X_T)]$ .

La  $m_t = \mathbb{E}[f(X_t)]$ . Da får vi ved å bruke Itô's formel med,  $Y(x) = f(x)$ ,  
 $\frac{\partial}{\partial x} f(x) = a(x - b)$  og  $\frac{\partial^2}{\partial x^2} f(x) = a$ :

$$\begin{aligned} dY(t) &= df(X_t) = \frac{\partial}{\partial x} f(X_t) dX_t + \frac{1}{2} \frac{\partial^2}{\partial x^2} f(X_t) (dX_t)^2 \\ &= a(X_t - b) dX_t + \frac{1}{2} a (dX_t)^2 \\ &= a(X_t - b) (-au(t)(X_t - b)dt + u(t)\sqrt{\eta\Sigma}dB_t) + \frac{1}{2} au(t)^2 \eta \Sigma dt \end{aligned}$$

$$\begin{aligned}
&= \left( -2au(t) \frac{1}{2} a(X_t + b)^2 + \frac{1}{2} au(t)^2 \eta \Sigma \right) dt + a(X_t + b)u(t) \sqrt{\eta \Sigma} dB_t \\
&= \left( -2au(t) f(X_t) + \frac{1}{2} au(t)^2 \eta \Sigma \right) dt + a(X_t + b)u(t) \sqrt{\eta \Sigma} dB_t
\end{aligned}$$

Integrerer så med hensyn på t:

$$f(X_t) = f(x_0) - \int_0^t (2au(s)f(X_s) + \frac{1}{2} au(s)^2 \eta \Sigma) ds + \int_0^t a(X_s + b)u(s) \sqrt{\eta \Sigma} dB_s$$

Tar vi nå forventningen på begge sider får vi:

$$\begin{aligned}
\mathbb{E}[f(X_t)] &= \mathbb{E}[f(x_0)] + \mathbb{E} \left[ \int_0^t \left( -2au(s)f(X_s) + \frac{1}{2} au(s)^2 \eta \Sigma \right) ds \right] + \\
&\quad \mathbb{E} \left[ \int_0^t a(X_s + b)u(s) \sqrt{\eta \Sigma} dB_s \right] \\
&= f(x_0) + \mathbb{E} \left[ \int_0^t \left( -2au(s)f(X_s) + \frac{1}{2} au(s)^2 \eta \Sigma \right) ds \right] \\
&= f(x_0) + \int_0^t \mathbb{E} \left[ -2au(s)f(X_s) + \frac{1}{2} au(s)^2 \eta \Sigma \right] ds \\
&= f(x_0) + \int_0^t \left( -2au(s) \mathbb{E}[f(X_s)] + \frac{1}{2} au(s)^2 \eta \Sigma \right) ds \quad \implies \\
\frac{d}{dt} \mathbb{E}[f(X_t)] &= -2au(t) \mathbb{E}[f(X_t)] + \frac{1}{2} au(t)^2 \eta \Sigma \\
\dot{m}(t) &= -2au(t)m(t) + \frac{1}{2} a \eta \Sigma u(t)^2
\end{aligned}$$

I første overgang bruker vi lineariteten til forventningen og martingale-egenskapen til Itô-integralet. I den siste overgangen har vi igjen derivert med hensyn på t som har gitt oss en ordinær differensialligning med  $m(t) = \mathbb{E}[f(X_t)]$ .

## 4.1 Løsning kontroll problem

Vi skal nå bruke Hamilton-Jacobi-ligningen og dynamisk programmering for å løse problemet og slik minimerer  $m(T)$ . Vi har ingen løpende avkastning så vi får en enkel avkastningsfunksjonal vi ønsker å maksimere:

$$P[\alpha(\cdot)] = g(x(T)) = -m(T)$$

Kontrollen er  $u : [0, T] \rightarrow [0, 1]$  og vi har et dynamisk system:

$$\begin{cases} \dot{m}(t) &= -2au(t)m(t) + \frac{1}{2} a \eta \Sigma u(t)^2 & (0 < t \leq T) \\ m_0 &= a(x_0 - b)^2 \end{cases} \quad (66)$$

Vi kan i dette tilfellet finne løsningen til det dynamiske systemet for alle  $u(\cdot)$  og  $t$ :

$$\begin{aligned} \dot{m}(t) &= -2au_t m(t) + \frac{1}{2}a\eta\Sigma u(t)^2 \\ \dot{m}(t) + 2au_t m(t) &= \frac{1}{2}a\eta\Sigma u(t)^2 \\ [e^{\int_0^t 2au(s)ds} m(t)]' &= e^{\int_0^t 2au(s)ds} \frac{1}{2}a\eta\Sigma u_t^2 \\ e^{\int_0^t 2au(s)ds} m(t) &= m_0 + \frac{1}{2}a\eta \int_0^t e^{\int_0^s 2au(\tau)d\tau} \Sigma u(s)^2 ds \\ m(t) &= m_0 e^{-\int_0^t 2au(s)ds} + \frac{1}{2}a\eta\Sigma e^{-\int_0^t 2au(s)ds} \int_0^t e^{\int_0^s 2au(\tau)d\tau} u(s)^2 ds \\ m(t) &= m_0 e^{-\int_0^t 2au(s)ds} + \frac{1}{2}a\eta\Sigma \int_0^t e^{-\int_s^t 2au(\tau)d\tau} \Sigma u(s)^2 ds \end{aligned}$$

Verdifunksjonen blir i dette tilfellet:

$$v(t, m) = \min_{u(t) \in \mathcal{A}} P_{t,m}[u(t)] = \min_{u(t) \in \mathcal{A}} m(T)$$

Vi kan og vise at  $\partial_m v(t, m) > 0$  for alle  $a$ :

$$\begin{aligned} \frac{\partial}{\partial m} v(t, m) &= \frac{v(t, m+h) - v(t, m)}{h} \\ &= \frac{1}{h} \left[ (m_0 + h) e^{-\int_0^T 2au(s)ds} + \frac{1}{2}a\eta\Sigma \int_0^T -e^{\int_s^T 2au(\tau)d\tau} \Sigma u(s)^2 ds \right. \\ &\quad \left. - m_0 e^{-\int_0^T 2au(s)ds} - \frac{1}{2}a\eta\Sigma \int_0^T -e^{\int_s^T 2au(\tau)d\tau} \Sigma u(s)^2 ds \right] \\ &= \frac{1}{h} \left[ (m_0 + h) e^{-\int_0^T 2au(s)ds} - m_0 e^{-\int_0^T 2au(s)ds} \right] \\ &= e^{-\int_0^T 2au(s)ds} \geq 0, \quad \forall a \text{ og } t \in [0, T] \end{aligned}$$

Hamilton-Jacobi-ligningen (36) for vår dynamikk og kontroll er:

$$\partial_t v(t, m) + \min_{u:[t,T] \rightarrow [0,1]} \left\{ \partial_m v(t, m) \left( -2aum + \frac{1}{2}a\eta\Sigma u^2 \right) \right\} = 0$$

Vi har allerede vist at  $\partial_m v(t, m) > 0$  så vi kan flytte ut  $\partial_m v(t, m)$  og står igjen med:

$$\min_{u:[t,T] \rightarrow [0,1]} \left\{ \left( -2aum + \frac{1}{2}a\eta\Sigma u^2 \right) \right\}$$

$u$ ,  $\eta$  og  $\Sigma$  er alle positive for alle  $t$  så vi står igjen med de to tilfellene  $a > 0$  og  $a < 0$ .

Gitt  $a < 0$  er løsningen av det dynamiske systemet:

$$m_t = a(x_0 - b)^2 e^{\int_0^t 2ku(s)ds} + \frac{a}{2}\eta\Sigma \int_0^t e^{\int_s^t 2ku(\tau)d\tau} \Sigma u(s)^2 ds$$

Begge leddene er negative for alle  $t, u, b$  og  $x_0$ . Så i tilfellet  $a < 0$  så er og  $m_t < 0$ . Vi har da minimeringsproblemet for  $a < 0$ :

$$\min_{u: [t, T] \rightarrow [0, 1]} \left\{ -a(2um - \frac{1}{2}\eta\Sigma u^2) \right\} = (-a) \min_{u: [t, T] \rightarrow [0, 1]} \left\{ f(u) \right\}$$

Hvor  $f(u) = 2um - \frac{1}{2}\eta\Sigma u^2$ .  $f(u)$  er konkav i  $u$  og ettersom  $m < 0$  så oppnår vi maksimum for  $f(u)$  når  $u = \frac{2m}{\eta\Sigma} < 0$ , siden  $\eta$  og  $\Sigma$  er positive og  $m$  negativ. Vi velger da ekstremalpunktet  $u = 1$  for å minimere  $f(u)$ .

Velger vi  $a > 0$  er  $m \geq 0$  og er  $f(u) = -2mu + \frac{1}{2}\eta\Sigma u^2$  og vi finner minimum når  $f'(u) = -2m + \eta\Sigma u = 0$ . Dette gjør at vi velger  $u = \frac{2m}{\eta\Sigma} \geq 0$  så lenge  $\frac{2m}{\eta\Sigma} \leq 1$ . Ellers velger vi  $u = 1$ .

Dette gir oss en strategi for valg av  $u(\cdot)$  ved alle  $t$ :

$$u(t) = \begin{cases} 1 & \text{Hvis } a \leq 0 \\ \min(1, \frac{2m}{\eta\Sigma}) & \text{Hvis } a > 0 \end{cases} \quad (67)$$

Setter vi dette tilbake i løsningen  $m(t)$  for  $a \leq 0$  får vi:

$$\begin{aligned} m(t) &= m_0 e^{-\int_0^t 2a_1 ds} + \frac{1}{2} a \eta \Sigma \int_0^t e^{-\int_s^t 2a_1 d\tau} 1 ds \\ &= m_0 e^{-2at} + \frac{1}{2} a \eta \Sigma \int_0^t e^{-2a(t-s)} ds \\ &= m_0 e^{-2at} + \frac{1}{2} a \eta \Sigma \left[ e^{-2at} \frac{e^{2as}}{2a} \right]_0^t = m_0 e^{-2at} + \frac{1}{4} \eta \Sigma [1 - e^{-2at}] \end{aligned}$$

For  $a > 0$  har vi samme løsning som  $a < 0$  helt frem til punktet  $t^*$ , hvor  $\frac{2m(t^*)}{\eta\Sigma} = 1$ . Dette gir  $m(t^*) = \frac{\eta\Sigma}{2}$  ettersom vi velger  $u(t^*) = 1$  for dette punktet.

Vi setter  $u(t) = \frac{2m}{\eta\Sigma}$  inn i (66) og kan forenkle løsning for  $m$ :

$$\begin{aligned} \dot{m}(t) &= -2a \frac{2m(t)}{\eta\Sigma} m(t) + a \frac{2m(t)^2}{\eta\Sigma} = -\frac{2a}{\eta\Sigma} m(t)^2 \implies \\ \frac{\dot{m}(t)}{m(t)^2} &= -\frac{2a}{\eta\Sigma} \end{aligned}$$

Dette er en separabel ligning så vi løser ved å integrere med hensyn på  $t$  på begge sider.

$$\int_{t^*}^t \frac{\dot{m}(s)}{m(s)^2} ds = -\frac{2a}{\eta\Sigma} (t - t^*)$$

For å løse venstre side bruker vi delvis integrasjon:

$$\begin{aligned} \int_{t^*}^t \frac{\dot{m}(s)}{m(s)^2} ds &= \frac{1}{m(s)} \Big|_{t^*}^t - \int_{t^*}^t m(s) \left( \frac{1}{m(s)^2} \right)' ds = \frac{1}{m(s)} \Big|_{t^*}^t - \int_{t^*}^t m(s) \left( \frac{-2}{m(s)^3} \dot{m}(s) \right) ds \\ &= \frac{1}{m(s)} \Big|_{t^*}^t + 2 \int_{t^*}^t \frac{\dot{m}(s)}{m(s)^2} ds \\ \implies \int_{t^*}^t \frac{\dot{m}(s)}{m(s)^2} ds &= -\frac{1}{m(s)} \Big|_{t^*}^t = \frac{1}{m(t^*)} - \frac{1}{m(t)} = -\frac{2a}{\eta\Sigma} (t - t^*) \end{aligned}$$

Videre bruker vi at  $m(t^*) = \frac{\eta\Sigma}{2}$  og rydder opp i ligningen og får uttrykk for  $m(t)$  for  $a > 0$  og  $t > t^*$ :

$$m(t) = \frac{\eta\Sigma}{2 + 2a(t - t^*)}$$

så vi kan summere opp:

$$m(t) = \begin{cases} m_0 e^{-2at} + \frac{1}{4}\eta\Sigma[1 - e^{-2at}] & \text{Når } a < 0 \text{ og } t \leq t^* \\ \frac{\eta\Sigma}{2+2a(t-t^*)} & \text{Når } a > 0 \text{ og } t > t^* \end{cases} \quad (68)$$

Det siste vi trenger nå er å finne  $t^*$ . Ved  $t^*$  er  $u(t^*) = 1$ , og  $u(t^*) = \frac{\eta\Sigma}{2m(t^*)} \implies 2m(t^*) = \eta\Sigma$ :

$$\begin{aligned} m(t^*) &= m_0 e^{-2at^*} + \frac{1}{4}\eta\Sigma[1 - e^{-2at^*}] \\ \frac{2m(t^*)}{\eta\Sigma} &= \frac{2m_0 e^{-2at^*}}{\eta\Sigma} + \frac{1}{2}[1 - e^{-2at^*}] = \frac{2m_0 e^{-2at^*}}{\eta\Sigma} + \frac{1}{2} - \frac{1}{2}e^{-2at^*} = 1 \implies \\ \frac{1}{2} &= e^{-2at^*} \left( \frac{2m_0}{\eta\Sigma} - \frac{1}{2} \right) \quad \text{Så tar vi logaritmen på begge sider:} \end{aligned}$$

$$\begin{aligned} \ln\left(\frac{1}{2}\right) &= -\ln(2) = \ln\left[e^{-2at^*} \left( \frac{2m_0}{\eta\Sigma} - \frac{1}{2} \right)\right] \\ &= \ln(e^{-2at^*}) + \ln\left[\frac{2m_0}{\eta\Sigma} - \frac{1}{2}\right] = -2at^* + \ln\left[\frac{2m_0}{\eta\Sigma} - \frac{1}{2}\right] \implies \\ 2at^* &= \ln(2) + \ln\left[\frac{2m_0}{\eta\Sigma} - \frac{1}{2}\right] \end{aligned}$$

som gir oss om vi rydder opp får vi  $t^*$ :

$$t^* = \frac{1}{2a} \left[ \ln(2) + \ln\left[\frac{2m_0}{\eta\Sigma} - \frac{1}{2}\right] \right] = \frac{1}{2a} \ln\left[\frac{4m_0}{\eta\Sigma} - 1\right]$$

Vi har altså nå funnet en optimal strategi for å minimere forventningen til f.

$$u(t) = \begin{cases} 1 & \text{Hvis } a < 0 \text{ og } t \leq t^* \\ \frac{1}{2+2a(t-t^*)} & \text{Hvis } a > 0 \text{ og } t > t^* \end{cases}$$

## 4.2 Overgang til $\mathbb{R}^d$ med eksempel

Vi har foregående gått gjennom konstruksjon av en kontroll for tilfellet i  $\mathbb{R}$ . Vi har i utgangspunktet et minimeringsproblemet i  $\mathbb{R}^d$ :

$$\min_{x \in \mathbb{R}^d} f(x) = \min_{x \in \mathbb{R}^d} \frac{1}{n} \sum_{i=1}^n f_i(x)$$

For å kunne bruke det vi har kommet frem til allerede skal vi bruke en lokal diagonal kvadratisk tilnærming. Vi antar nå at  $f(x) \approx \frac{1}{2} \sum_{i=1}^d a_i (x_i - b_i)^2$  lokalt

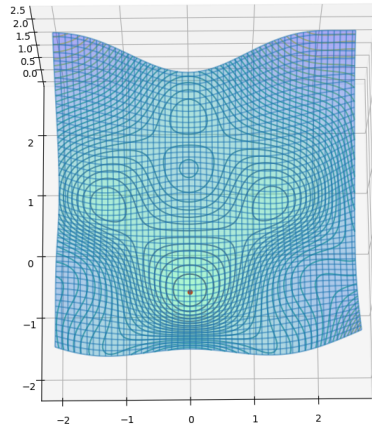


i  $x$ , hvor  $a_i, b_i \in \mathbb{R}$ . Videre antar vi at  $\Sigma(x) \approx \text{diag}\{\Sigma_1, \Sigma_2, \dots, \Sigma_d\}$  hvor  $\Sigma_i$  er konstant lokalt for alle  $i$ . Vi kan nå se på læringsrate for hver enkelt dimensjon og innføre kontroll  $u_i$  tilsvarende. Ettersom vi kun approksimerer lokalt må vi estimere  $a_i, b_i$  kontinuerlig. Det finnes her flere alternativer og vi skal ikke gå inn i detaljer. I stedet skal vi vise et eksempel på en implementering av denne alternative stokastiske gradientalgoritmen.

#### 4.2.1 Implementering

Vi skal nå implementere algoritmen på funksjonen:

$$f(x, y) = \frac{\sin(2y)\cos(2x)}{2} + \frac{x^2 + y^2}{5} + 0.4$$



Figur 3: Oversikt over funksjonen plottet med minimumspunkt  $(0.0, -0.65322)$ .

Det første vi gjør er å skrive om  $f$ :

$$\begin{aligned} f(x, y) &= \frac{\sin(2y)\cos(2x)}{2} + \frac{(x-1)^2 + (x+1)^2 - 2}{10} + \frac{(y-1)^2 + (y+1)^2 - 2}{10} + 0.4 \\ &= \frac{\sin(2y)\cos(2x)}{2} + \frac{(x-1)^2}{10} + \frac{(x+1)^2}{10} + \frac{(y-1)^2}{10} + \frac{(y+1)^2}{10} \end{aligned}$$

Vi kan nå finne de partiellderiverte til  $f$ :

$$\begin{cases} \partial_x f(x, y) &= -\sin(2y)\sin(2x) + \frac{x-1}{5} + \frac{x+1}{5} \\ \partial_y f(x, y) &= \cos(2y)\cos(2x) + \frac{y-1}{5} + \frac{y+1}{5} \end{cases} \quad (69)$$

Omskrivningen gjør at vi får  $\nabla f(x, y)$  som 6 funksjoner.  $f_j$  er en av de tre funksjonene i  $\partial_y f(x, y)$  og  $f_i$  er en av funksjonene i  $\partial_x f(x, y)$ . Vi bruker så disse 6 funksjonene til å konstruere to ulike typer  $\Sigma(x, y)$ . Den første er:

$$\Sigma_1(x, y) = \frac{1}{6} \left( \sum_{i=1}^3 \left( \partial_x f(x, y) - f_i(x, y) \right)^2 + \sum_{j=1}^3 \left( \partial_y f(x, y) - f_j(x, y) \right)^2 \right)$$

Og den andre:

$$\Sigma_2(x, y) = \sum_{i=1}^3 f_i(x, y)^2 + \sum_{j=1}^3 f_j(x, y)^2$$

Begge disse er positive for alle  $x$  og  $y$ .

Vi velger så et startpunkt  $(x_0, y_0)$  og en læringsrate  $\eta > 0$ . Da blir stegene i algoritmen for iterasjon nummer  $j$  som følgende.

**Steg 1:** For iterasjon  $(k+1)$  regner vi ut  $\Sigma(x_k, y_k)$  og trekker en tilfeldig fra de partiellderiverte. Vi antar at vi plukker en  $f_j(x, y)$  fra  $\partial_x f$ .

Vi skal for hver iterasjon finne en funksjon  $g$  slik at  $g(x) = \frac{a}{2}(x - b)^2 \approx f(x, y)$  ved å regner ut tilhørende  $a$  og  $b$ . Vi bruker at  $\frac{d}{dx}g(x_k) = a(x - b) = f_j(x_k, y_k)$  og at  $g(x_k) = f(x_k, y_k)$  for iterasjon  $k$  da blir:

$$b = x_k - 2 \frac{f(x_k, y_k)}{f_j(x_k, y_k)}$$

$$a = \frac{f_j(x_k, y_k)^2}{2f(x_k, y_k)}$$

og med  $a$  og  $b$  regner vi ut:

$$m_k = a(x_k - b)^2$$

**Steg 2:** Avgjør om  $a > 0$ . Dersom  $a < 0$  lar vi læringsraten være og oppdaterer med følgende skjema og starter en ny iterasjon:

$$\begin{cases} x_{k+1} = x_k + \eta f_j(x_k, y_k) \\ y_{k+1} = y_k \end{cases}$$

Vi oppdaterer  $y$  i stedet for  $x$  i de tilfellene hvor vi plukker fra  $\partial_y f$ .

**Steg 3:** Dersom  $a > 0$  sjekker vi om  $\frac{2m_k}{\eta \Sigma(x_k, y_k)} \geq 1$ . Om det stemmer følger vi skjema i steg 2. I tilfellet der  $\frac{2m_k}{\eta \Sigma(x_k, y_k)} < 1$  oppdaterer vi etter skjema:

$$\begin{cases} x_{k+1} = x_i + u_k \eta f_j(x_k, y_k) \\ y_{k+1} = y_k \end{cases} \quad \text{Hvor} \quad u_k = \frac{2m_k}{\eta \Sigma(x_k, y_k)}$$

Før vi går videre til neste iterasjon.

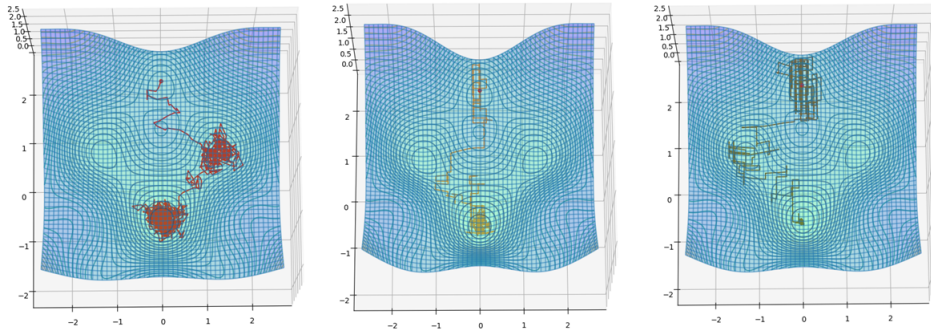
#### 4.2.2 Resultater av test

I denne testen skal vi sammenligne 3 ulike algoritmer med hverandre. De baserer seg alle 3 på inndelingen av  $f(x, y)$  som gjort over og de partiellderiverte (69). Algoritme **(1)** er uten kontroll. **(2)** og **(3)** er begge med kontroll, men i **(2)** har vi brukt  $\Sigma_1$  og i **(3)** har vi brukt  $\Sigma_2$ .

Testen er utført med 500 tilfeldig valgte startpunkter innefor  $(-2, 2) \times (-2, 2)$ . Det er så regnet ut et snitt på avstand mellom de siste punktet  $k$  i algoritmen

og minimum til funksjonen:  $|f(x_k, y_k) - f(0, 0.65322)|$ . Det er og testet med 3 ulike læringsrater og iterasjoner.

Figuren til venstre under viser en sti for en stokastisk gradientalgoritme hvor vi har fulgt  $\eta \nabla f(x, y) + \eta \xi$ , hvor  $\xi$  er tilfeldig støy lagt til i hver iterasjon. Denne tilnærmingen gir enda dårligere resultater enn alternativene over og er derfor utelatt fra resten av testen.



Figur 4: plot av stier for ulike tilnærmingene til SGD. Til høyre er sti til algoritme **(3)** og i midten algoritme **(1)**

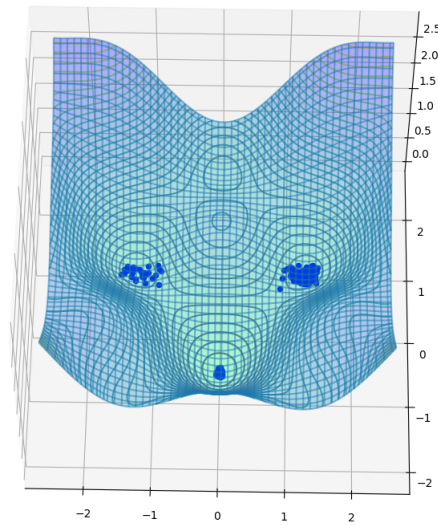
Vi starter med å se på utgangspunktet vårt. den stokastiske gradientalgoritme hvor vi har delt gradienten inn i de 6 funksjoner som over(69):

$$x_{k+1} = x_k - \eta \nabla f_{\delta(k+1)}(x_k)$$

antall startpunkter	læringsrate	iterasjoner	Snitt resultat
500	1	300	0.2365
500	1	500	0.1985
500	1	1000	0.1687
500	3	300	0.1855
500	3	500	0.1790
500	3	1000	0.1474
500	5	300	0.2115
500	5	500	0.1955
500	5	1000	0.1655

Tabell for standard stokastisk gradientalgoritme **(1)**

Som vi kan lese ut av tabellen for stokastisk gradientalgoritme er det læringsrate på 3 som gir best resultater og øker vi antallet iterasjoner blir og algoritmen bedre. Ved å øke antall iterasjoner for  $\eta = 3$  gradvis til 5000 kan vi forbedre snittresultatet, men det flater ut rundt 0.09 i denne testen.



Figur 5: Illustrasjon av 300 endepunkt for algoritme **(1)** med  $\eta = 3 \cdot 5000$  iterasjoner.

I plottet av endepunktene til algoritme **(1)**, i figur 1 ser vi at algoritmen ikke er så pålitelig. Vi kan ikke være sikker på at vi er kommet frem til riktig dal.

Tabell for algoritme **(2)**:

antall startpunkter	læringsrate	iterasjoner	Snitt resultat
500	1	300	0.0239
500	1	500	0.0118
500	1	1000	0.0089
500	3	300	0.0202
500	3	500	0.0114
500	3	1000	0.0092
500	5	300	0.0298
500	5	500	0.0116
500	5	1000	0.0090

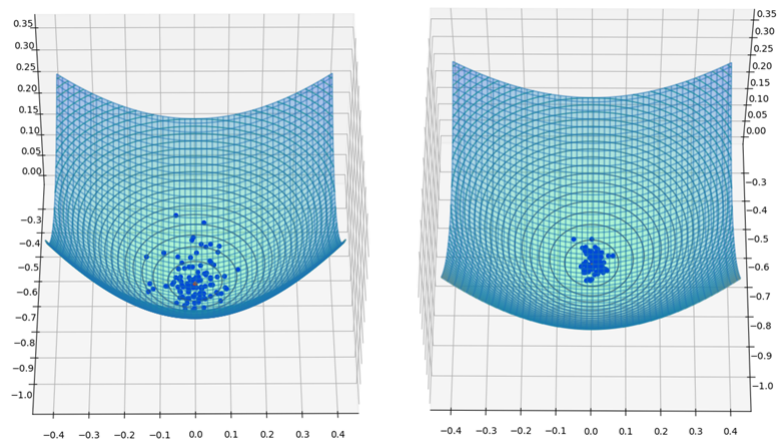
Tabell for tilpasset stokastisk gradientalgoritme **(2)**

Tabell for algoritme **(3)**:

antall startpunkter	læringsrate	iterasjoner	Snitt resultat
500	1	300	0.0087
500	1	500	0.006
500	1	1000	0.0006
500	3	300	0.0014
500	3	500	0.0005
500	3	1000	0.0005
500	5	300	0.0006
500	5	500	0.0005
500	5	1000	0.0005

Tabell for tilpasset stokastisk gradientalgoritme **(3)**

Vi kan lese ut av tabellene til de to tilpassede algoritmene er at prestasjonen er vesentlig bedre enn standard stokastisk gradientalgoritme **(1)**. Dette tyder på at kontrollen i algoritmene **(2)** og **(3)** fungerer slik vi ønsker. Vi kan og se at algoritmene ikke er like følsomme for valg av læringsrate, etter som det i begge tilfeller, når antall iterasjoner overstiger 300, så er resultatene så og si like for alle læringsratene.



Figur 6: 300 endepunkter til algoritme **(2)** og **(3)**

Vi kan og se ut fra plottet 6 av endepunktene at begge algoritmene ender opp rundt det faktiske minimum av funksjonen som igjen bekrefter at algoritmene fungerer bedre enn for algoritme **(1)**.

## 5 Oppsummering

Vi har i denne oppgaven vist at vi den stokastiske gradientalgoritmen konvergerer mot en løsning av en stokastisk differensialligning når læringsraten  $\eta \rightarrow 0$ . Dette resultatet har vi så bygget videre på og vist at vi kan ved bruk av stokastisk analyse og optimal kontrollteori konstruere en alternativ stokastisk gradientalgoritmen hvor vi kontrollerer læringsraten i algoritmen. Algoritmen har vi så testet på et konkret eksempel hvor vi viser at algoritmen gir et betydelig bedre resultat enn den standard stokastisk gradientalgoritmen.

## 6 Referanser

- [1] Bahouri, H., Chemin, J. & Danchin, R. (2011). *Fourier analysis and non-linear partial differential equations* Springer
- [2] Bach, F. (2020, 1. mai). *Effortless optimization through gradient flows*. Machine Learning Research Blog. <https://francisbach.com/gradient-flows/>
- [3] Bellman, R. (1957). *Dynamic programming*. Princeton university press.
- [4] Evans, L. C., (u.å). *An Introduction to Mathematical Optimal Control Theory* forelesningsnotater fra Universitetet i Carlifornia, Berkeley, url: <https://math.berkeley.edu/~evans/control.course.pdf>
- [5] Evans, L. C., (2013). *An introduction to stochastic differential equations version 1.2* American Mathematical Society.
- [6] Fleming, W. H. & Rishel, R. W. (1975). *Deterministic and Stochastic Optimal Control*. Springer-Verlag.
- [7] Karatzas, I. & Shreve, S. (1991). *Brownian Motion and Stochastic Calculus(2. utg.)*. Springer-Verlag.
- [8] Lanconelli, A. & Lauria, C. S,(2022). A note on diffusion limits for stochastic gradient descent. <https://doi.org/10.48550/arXiv.2210.11257>.
- [9] Li, Q. , Tai, C. & E,W. (2017). Stochastic modified equations and the dynamics of stochastic gradient algorithms. *International Conference on Machine Learning*, (70), 2101-2110. <https://doi.org/10.48550/arXiv.1511.06251>
- [10] Stroock, D. & Varadhan, S. (2006). *Multidimensional Diffusion Processes*. Springer
- [11] Varadhan, S.R.S.,(2012) *Probability Theory (2.utg.)*. American Mathematical Society.
- [12] Øksendal, B. (2013) *Stochastic Differential Equations: An introduction with applications(6. utg.)*. Springer.