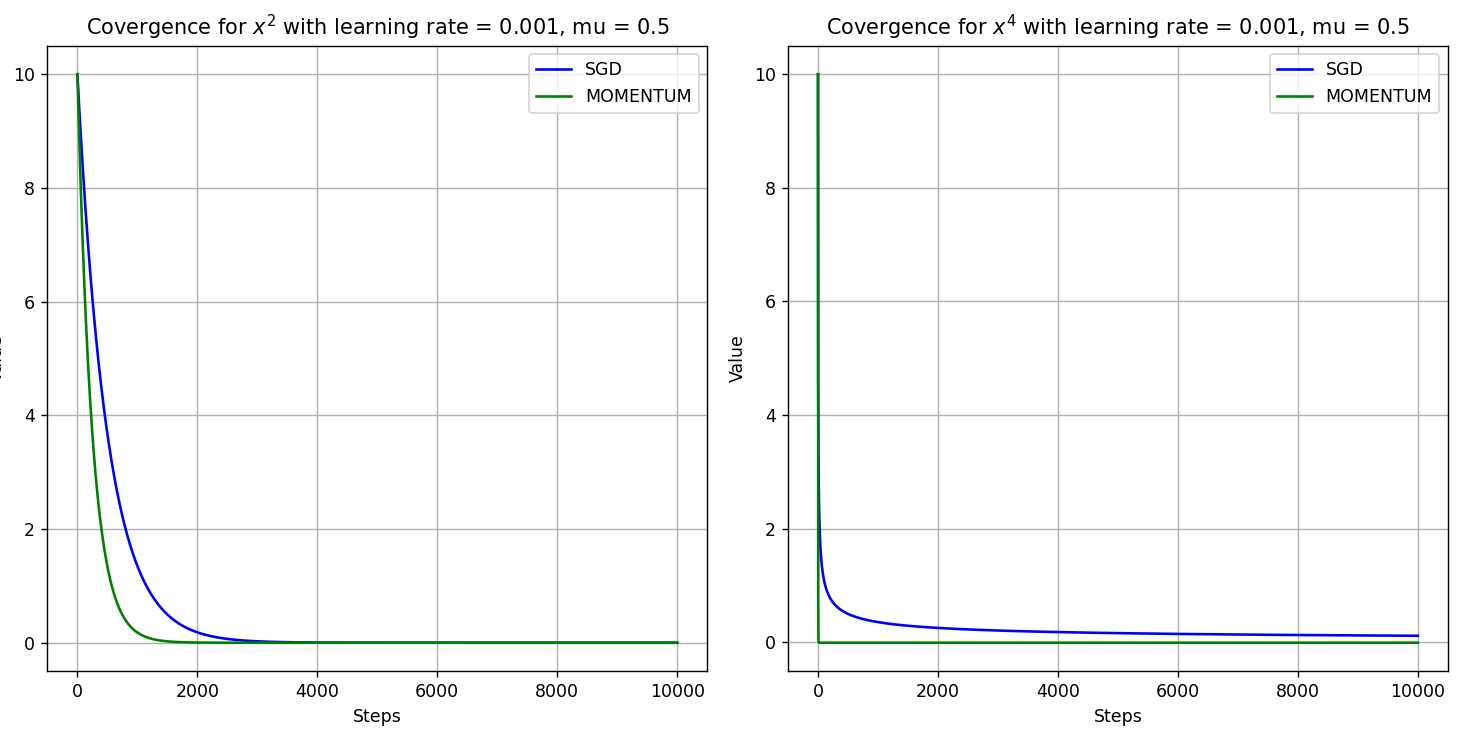
בס"ד

**מגיש:** מיכאל לוגסי **ת.ז:** 305536575

**דו"ח תרגיל בית 3**

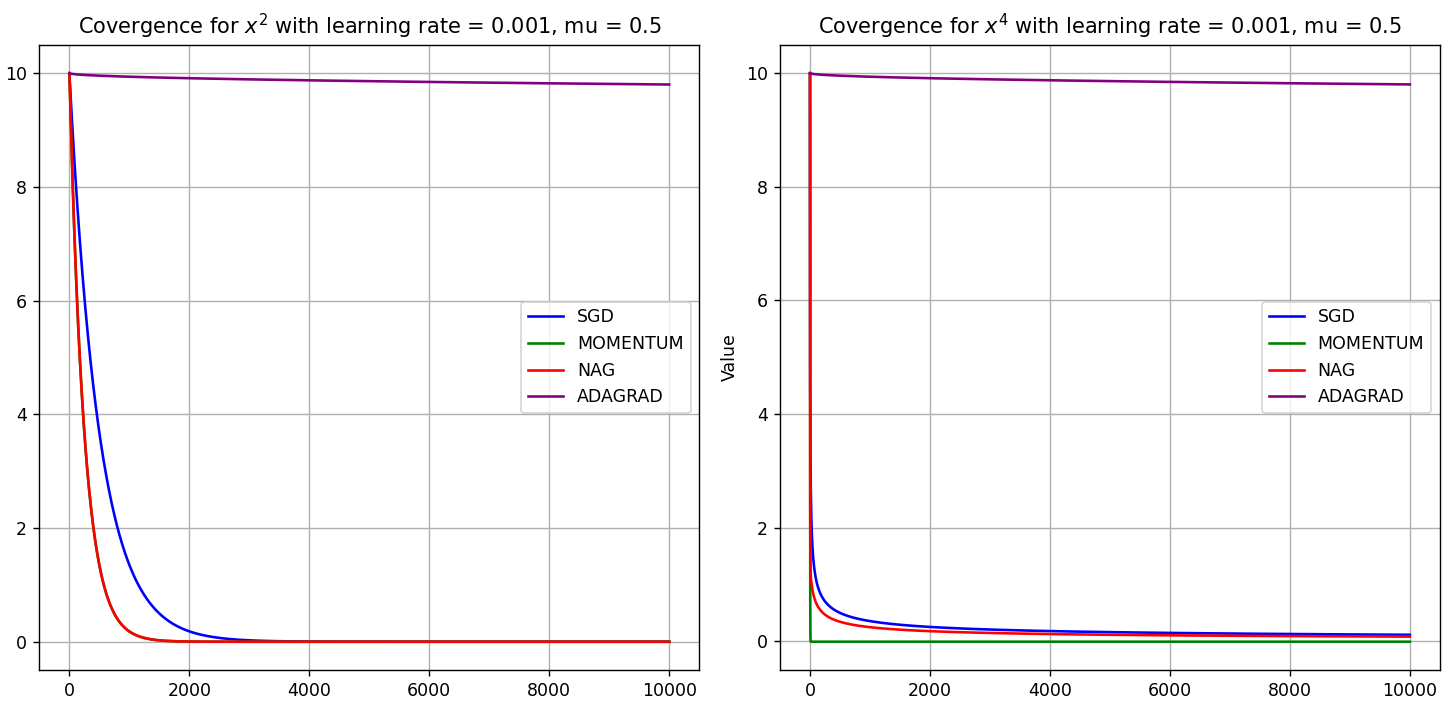
תיאור העבודה:  
קיבלנו קוד למימוש של 4 פונקציות שונות למציאת מינימום של שתי פונקציות באמצעות גארדיאנט. מימשתי את הפונקציות הנדרשות תוך כדי התאמות לפנוקציות שהיו בטמפלייט והוספת משתנים.  
לאחר מכן ביצעתי בדיקות לריצה ולתוצאות ע"מ לראות את ההתכנסות של כל פונקציה למציאת המינימום תוך התאמה של ה learning rate ו momentum כדי לראות תוצאות מספיק טובות ויציבות עבור כל אחת מהשיטות.  
למדתי מההתנסות הזאת שבעצם אין דבר כזה שיטה הכי טובה אלא שיטה שטובה בדר"כ ושעבור כל מקרה ניתן אולי למצוא פונקציה טובה יותר מזו שבדר"כ ושגם מאוד חשוב להתאים את קצב הלימוד והמומנטום שיהיו טובים כדי לראות תוצאות מספקות עבור כל אחת מהשיטות.

1. A graph of a graph

   Description automatically generated with medium confidenceמימוש בוצע בקובץ הפיתון – בתחילה הרצתי עם learning rate של 0.1 וניתן היה ליראות שעבור הפונקציה של x^2 הגראדיאנט הגיע במהירות ל 0, אך עבור הפונקציה X^4, הגראדיאנט לא עבד כלל, זה הגיע למצב שהריצה קרסה מאחר ומהספרים היו גדולים או קטנים (לכיוון השלילי) מידי.  
   שיחקתי קצת עם ה learning rate וראיתי שעבור learning rate של 0.001 הגארדיאנט מתכנס עבור שתי הפונקציות, בגרף למטה ניתן לראות שעבור X^4 הגארדיאנט מתקדם מהר יותר מאחר ותוצאת הנגזרת שלו גדולה יותר ולכן הקצב לימוד גדול יותר, אך ניתן לראות שעבור ההפוקנציה X^2 הגארדיאנט אמנם התקדם לאט אך הגיע קרוב יותר ל 0, ככה שקצב הלימוד האיטי שלו עזר לו להתקב יותר למינימום.
2. A green and blue graph

   Description automatically generatedמימוש הפונקציה של המומנטום גם כן נמצאת בקוד פייתון, בתהחלה שמתי ממונטום כך שיהיה משלים ל 1 עם ה learning rate כמו שלמדנו בכיתה, כאשר בחרתי ללכת עם ה learning rete של 0.001 שראיתי שעובד עבור שתי הפונקציות. התוצאות הראו שמומנטום עבור שתי הפונקציות גרם לתנודטיות גדולה במקום לכנס את ה X למינימום, כך שנראה שנתונים שנתתי הגראדיאנט היה טוב יותר, מהיר ויציב יותר.  
     
   לכן שיחקתי קצת עם המומנטום וראיתי שכשאר אני בוחר מומנטום של 0.5 (כלומר ממוצע עם הגאדיאנט) התוצאות משתפרות עבורו (שוב עם אותו learning rate) ומראות שהוא מגיע מהר יותר וקרוב יותר למינימום.

לגבי השאלה – ניתן בהחלט לראות השפעה של המומנטום גם לחיוב וגם לשלילה עבור שתי הפונקציות. ראינו שכשיש לו ערך גבוהה מאוד הוא משפיע חזק מידי על ה X ככה שהוא מתבדר ומתקשה יותר להכנס אך כאשר מתאימים אותו ניתן לראות שהוא משפיע לחיוב על ההתקדמות וגרם לפונקציות להתכנס למינימום מהר הרבה יותר.

1. הוספתי את שתי השיטות הנוספות adagrad ו NAG, והרצתי עם אותם learning rate ומומנטום שעבדו לי טוב בשאלה הקודמת, learning rate של 0.001 ומומנטום של 0.5  
   ניתן לראות שה adagrad מתקדם מאוד מאוד לאט ולא מצליח להתכנס ל 0, כל השאר פחות או יותר די קרובים כאשר נראה שעבור המקרה שלנו ה momentum הוא המהיר ביותר עבור שתי הפונקציות.