Analysis of Projected Gradient Descent

Anton Xue and Nikolai Matni

1 Introduction

Projected Gradient Descent

2 Background

Stephen Boyd and Lieven Vandenberghe [1].

2.1 Quadratic Programs with Linear Equality

A quadratic program with equality constraints [1] is

$$\begin{aligned} \text{minimize} \quad & \frac{1}{2} x^\top M x & (1) \\ \text{subject to} \quad & Ax = b & (2) \end{aligned}$$

with variable in $x \in \mathbb{R}^n$, where $Q \succeq 0$ and the constraint $A \in \mathbb{R}^{m \times n}$ is, for our purposes, fat and full rank.

2.2 Projected Gradient Descent

Projected gradient descent has the step rule

$$x^+ = \hat{x} + (I - A^+ A)(I - tM)x$$

where A^+ is the Moore-Penrose pseudoinverse of A (such that $I - A^+ A$ projects onto $\ker A$); the least-norm solution of the linear constraint is $\hat{x} = A^+ b$; and $t < 1/L$ where $L = \lambda_{\max}(M)$ is the Lipschitz constant of the gradient.

3 Taking steps

We seek to characterize the space in which projected gradient descent traverses.

Starting at some initial x_0, note that

$$x_{k+1} = (I - A^+A)(I - tM)x_k + \hat{x} = \big[(I - A^+A)(I - tM)\big]^k x_0 + \sum_{l=0}^{k-1} \big[(I - A^+A)(I - tM)\big]^l \hat{x}$$

However since $t < 1/\lambda_{\max}(M)$, the matrix $I - tM$ is invertible; furthermore, since $I - A^+A$ is symmetric, the two commute (c.f., [2] Theorem 4.5.17(a)):

$$(I - A^+A)(I - tM) = (I - tM)(I - A^+A)$$

and because $I - A^+A$ is a projection matrix, $(I - A^+A)^l = I - A^+A$ for $l > 0$. By definition of $\hat{x}$, we also have that $(I - A^+A)\hat{x} = 0$. Thus, the recursion simplifies to

$$x_{k+1} = (I - A^+A)(I - tM)^k x_0 + \hat{x}$$

and in the limit, supposing that $W^\top W$ projects onto $\ker M$,

$$x_\infty = (I - A^+A)W^\top W x_0 + \hat{x}$$

References

[1] Stephen P Boyd and Lieven Vandenberghe. *Convex Optimization*. Cambridge University Press, 2004.

[2] Roger A Horn and Charles R Johnson. *Matrix Analysis*. Cambridge University Press, 2012.