

# Convex Optimization for Big Data

Mario Vázquez Corte 127252

May 2018

## 1 Introduction

El artículo es un survey sobre el estado y los avances de la optimización convexa enfocados en Big Data. El paper pone énfasis en el procesamiento de señales.

## 2 Optimización Convexa en la era de Big Data

Los autores mencionan que la optimización convexa ha tenido gran avance impulsado por su utilidad, y las necesidades creadas por el área de Big Data. Los algoritmos convencionales/antiguos no han sido capaces de resolver los nuevos problemas, pues la cantidad de datos es relativamente masiva.

## 3 "The Basics"

El paper describe un problema básico de optimización en Big Data:

$$F^* \stackrel{\text{def}}{=} \min_x \left\{ F(x) \stackrel{\text{def}}{=} f(x) + g(x) : x \in \mathbb{R}^p \right\}$$

Prosigue a estudiar los métodos de optimización más comunes para este tipo de problemas.

1. CPO
  - Poseen exactitud media/baja. Utilizan condiciones/información de primer orden.
2. Random
  - Permiten mayor escalabilidad.
3. Distribuidos/paralelos
  - Aprovechan las cualidades de los métodos CPO, pero de forma escalable.

### 3.1 CPO

Son los más comunes y utilizados. El paper describe el problema de minimos cuadrados, y su version LASSO que integra un metaparametro de penalización. El método más conocido para este tipo de problemás es descenso en gradiente, que tiene su implementación estocástica para Big Data.

Los autores prosiguen a analizar descenso en gradiente y sus variantes. En la primera variante se agrega un parametro que representa el "momentum", y se le conoce como gradiente acelerado. Este método esta inspirado en el modelado de péndulos o fluidos. Existe una variante de esto, cuando la función  $g$  presenta propiedades poco smooth, se le conoce como método acelerado de gradiente aproximado. Ambos métodos resultan eficientes. Finalmente los autores mencionan el algoritmo, ADDM que se utiliza para métodos en los que existen problemas adicionales.

### 3.2 "Big Data Scaling via Randomization"

Estos algoritmos no han sido muy exitosos para problemas de gran escala. El primero en ser estudiado es Descenso en Coordenadas. A diferencia de gradiente descendente, este algoritmo elige la coordenada más grande del gradiente y se mueve en esa dirección. Este método es computacionalmente barato, pero su orden de convergencia es más lento. Pasa a analizar gradiente estocástico, como ya se menciona, este método es una variante gradiente descendente que utiliza gradientes aproximados (generalmente se toma un subset de los datos y se evalúa el gradiente ahí). Este método a sido más exitoso. Finalmente, analizan los métodos de algebra lineal estocástica o "randomized". La idea es representar los problemas de manera "low-rank" o de dimensión/complejidad menor.

### 3.3 "Big Data Scaling via Randomization"