**Reporte de Optimización Convexa para Big Data**

En este paper se revisan y discuten el avance en el desarrollo algoritmos de optimización convexos para grandes volúmenes de datos a.k.a *big data* que tienen como objetivo reducir los cuellos de botella de almacenaje, computacionales y de comunicación

Se menciona que dichos algoritmos se basan en tres pilares: métodos de primer orden, escalamiento por aleatorización y cómputo en paralelo.

-Métodos de primer orden

Estos métodos utilizan las soluciones numéricas a partir de los métodos de optimización, como el gradiente, o incluso para funciones no-suaves mediante el proximal mapping.

También es mencionado que se pueden incluir regularizaciones, incluso a funciones no-suaves de manera muy eficiente, casi como para funciones que sí son suaves.

El método de gradiente calcula el gradiente de la función en un punto y avanza una cantidad determinada de pasos (learning rate) en la dirección don hay mayor decrecimiento o crecimiento (dependiendo del problema), posteriormente repitiendo el procedimento de manera iterativa.

El método de gradiente acelerado es un método óptimo de primer orden pues eśte tiene la mejor tasa de error del peor de los casos.

-Objetivo compuesto

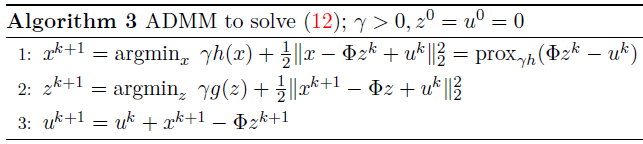
Aquí el paper discute el problema compuesto canónico (funcion objetivo convexa diferenciable conjuntada con una funcion no convexa)

Se discute que los métodos de descenso por gradiente aprovechan la estructura compuesta con el fin de qu las tasas de convergencia se mantengan para las clases de problemas suaves.

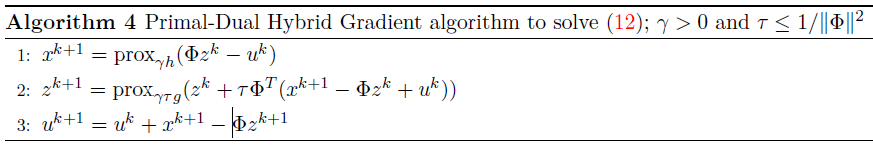
-Objetivos de proximidad

Se dice que los métodos de primer orden aun no son aplicables por que es de utilidad considerar una forma funcional como sigue:

Teniendo en cuenta que los operadores de proximidad de g y h, ambos eficientes, es posible aplicar un algoritmo simple, denominado método de multiplicación de multiplicadores (ADMM) para sus soluciones, que aprovecha poderosas técnicas de Lagrange y descomposición dual.



Sin embargo, se tienen algunos inconvenientes como que en ocasiones la matriz no es diagonalizable lo cual desembocaría en la inexistencia de la convergencia. Para solucionar lo anterior, el autor plantea el siguiente algoritmo primal dual híbrido.



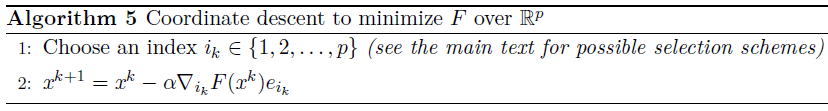
-Escalamiento Vía Aleatorización

Las técnicas de aleatorización mejoran la escalabilidad de los métodos de primer pues se controla su comportamiento de naturaleza esperado. Incluyen actualizaciones parciales aleatorias de variables de optimización, reemplazando el gradiente determinista y los cálculos proximales con estimadores estadísticos y acelerando las rutinas básicas de álgebra lineal mediante la aleatorización.

Aunque, en la práctica los cálculos numéricos y las iteraciones de los métodos de primer orden pueden hacer inviables estos métodos simples a medida del incremento en la dimensionalidad del problema. Se describen a continuación las aproximaciones aleatorias emergentes:

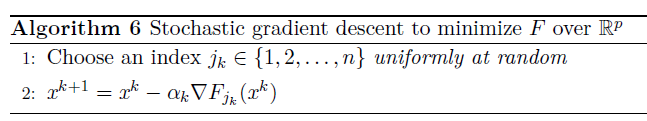
* Descenso Coordinado

El cálculo del gradiente en el problema de PageRank, requiere operaciones matriz-vector para cada iteración. Una mejor operación computacional sería elegir una coordenada i de x y únicamente modificar la variable correspondiente xi para mejorar la función objetivo.



* Descenso Estocástico

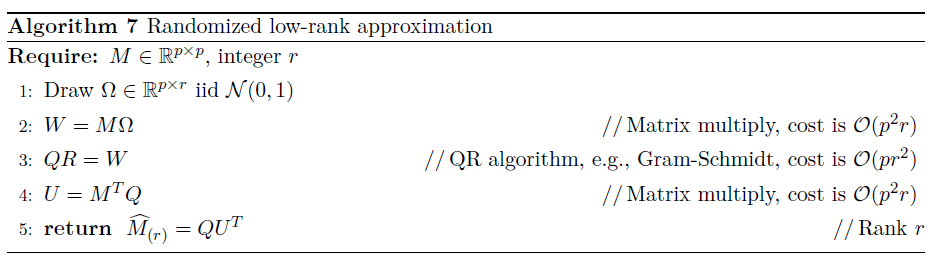
Este método actualiza todas las coordenas simultaneamente pero con el costo de usar gradientes aproximados, según lo siguiente:



* Álgebra lineal aleatorizada

Las operaciones de álgebra lineal básica pueden ser importantes cuellos de botella computacionales debido a su dependencia lineal de las dimensiones en el cotexto de grandes volúmenes de datos. Sin embargo, cuando los objetos de matriz relevantes tienen representaciones de rango bajo, la eficiencia de estos métodos mejora uniformemente de manera considerable.

La idea general es aproximar M= Q (QTM) con Q en , o para construir una representación de rango bajo por selección de subconjunto de columna o fila para acelerar el cálculo.



Cómputo en paralelo y distribuidos

Consisten en algoritmos paralelos sincrónicos idealizados con comunicaciones centralizadas a algoritmos asíncronos enormemente escalables con comunicaciones descentralizadas.

* Métodos de primer orden violentamente paralelos

Los métodos de primer orden pueden beneficiarse de la computación en paralelo. La expresión violentamente paralela se refiere a un escenario ideal para la paralelización en el que se divide el trabajo en cálculos independientes que se pueden realizar simultáneamente de manera predecible. Cada procesador se comunica con la ubicación central para formar el gradiente final y lograr la velocidad lineal ideal.

Los problemas de grandes volúmenes de datos requieren una revisión fundamental de cómo se diseñan los algoritmos de optimización convexos y se sugieren opciones computacionales no son muy convencionales.

Dado que las retricciones de sincronización y comunicación del hardware disponible naturalmente dictan la elección de los algoritmos, se espera que sigan descubriendo nuevas herramientas de aproximación que idealmente adapten los algoritmos convexos a la heterogeneidad de las plataformas computacionales. También se predice una mayor utilización de modelos compuestos y los correspondientes principios de mapeo proximal para problemas de *big data* no suaves para hacer frente al ruido y otras limitaciones.

Auque en general, me parece que en el futuro, se buscará la evolución de estas implementaciones o su conjunción con las técnicas actuales de escritura a disco como Hadoop o manejo en memoria como Spark.