Other types of methods for large-scale optimization

Computational Methods in Optimization
CS 520, Purdue

David F. Gleich
Purdue University
April 18, 2019

ALTERNATING OPTIMIZATION

Non-negative matrix factorization

$$\begin{aligned} \text{minimize} \quad & \|\mathbf{A} - \mathbf{X}\mathbf{Y}^T\| \\ \text{subject to} \quad & \mathbf{X} \in \mathbb{R}^{m \times k} \geq 0, \mathbf{Y} \in \mathbb{R}^{n \times k} \\ & \mathbf{X} \geq 0, \mathbf{Y} \geq 0 \end{aligned}$$

Fix X, solve for Y

Fix Y, solve for X

…

Does it converge?

Block coordinate descent

Gauss-Seidel

Alternating direction

Lots of activity among machine learning, compressed sensing, sparse 1-norm folks, too.

Bertsekas, Nonlinear programming

Suppose f is continuous, differentiable

$f = f(x_1, \ldots, x_N)$ where x_i is in a convex domain.
"Think of each x_i as a block of variables."

If

$$\underset{\mathbf{y} \in X_i}{\text{minimize}}\, f(\mathbf{x}_1, \ldots, \mathbf{y}, \ldots, \mathbf{x}_N)$$

is uniquely attained, then the sequence of subproblems converges to a stationary point.

Suppose there are just two blocks

[Grippo & Sciandrone]

Then we don't need a unique minimizer any more and we can treat more general convex problems.

STOCHASTIC GRADIENT DESCENT

SGD

Given $f(\mathbf{x}) = \sum_{i=1}^{L} f_i(\mathbf{x})$.

Note that $\mathbf{g}(\mathbf{x}) = \sum_{i=1}^{L} \nabla f_i(\mathbf{x})$

Consider $\mathbf{x}^{(k+1)} = \mathbf{x}^{(k)} - \alpha \nabla f_{i \sim U}(\mathbf{x})$

Here, $\nabla f_{i \sim U}(\mathbf{x})$ is just a random term in the gradient ("i drawn from uniform U")

Stochastic Gradient Descent

$$\text{minimize} \quad \|\mathbf{A}\mathbf{x} - \mathbf{b}\|^2$$

$$\text{minimize} \quad \sum_i \left(\sum_j A_{ij} x_j - b_i \right)^2$$

$$\text{minimize} \quad \textstyle\sum_i \ell_i(\mathbf{x})$$

$$\ell_i(\mathbf{x}) = \left(\sum_j A_{ij} x_j - b_i \right)^2$$

$$\mathbf{x}^{(k+1)} = \mathbf{x}^{(k)} - \alpha \mathbf{g}_{\ell_i}(\mathbf{x}^{(k)})$$

Repeatedly draw *i* at random.

$$= \mathbf{x}^{(k)} - \alpha 2 \left(\sum_j A_{ij} x_j - b_i\right) \begin{bmatrix} A_{i,1} \\ \vdots \\ A_{i,n} \end{bmatrix}$$