Multiple Gradient Descent Algorithm

Outi Wilppu

March 1, 2013

Motivation

Many real life problems have multiobjective nature

- Problem has several goals such that they should all be as good as possible
- Goals are conflicting and have to make compromises
- For example in economics and engineering
- Many problems have also non-smooth nature

Problem

Definition

We consider a multiobjective problem of form

$$\begin{aligned} \min \quad & f_1(\boldsymbol{x}), \ldots, f_k(\boldsymbol{x}) \\ \text{s. t.} \quad & \boldsymbol{x} \in S \end{aligned} \tag{1}$$

where

- Objective functions $f_i(\boldsymbol{x}) : \mathbb{R}^n \longrightarrow \mathbb{R}$
- Set $S \subseteq \mathbb{R}^n$ is a set of feasible solutions

Pareto optimality

Definition

A solution $\boldsymbol{x}^* \in S$ is Pareto optimal if there is no other solution $\boldsymbol{x} \in S$ such that $f_i(\boldsymbol{x}) \leq f_i(\boldsymbol{x}^*)$ for all $i = 1, \dots, k$ and $f_i(\boldsymbol{x}) < f_i(\boldsymbol{x}^*)$ for at least one index i.

- Usually there exists a lot of Pareto optimal solutions, called Pareto optimal set
- All Pareto optimal solution are mathematically equally good

Steepest Descent Method

- A method to solve smooth unconstrained optimization problem with single objective function

$$\min \ f(\boldsymbol{x}), \quad \text{s. t. } \ \boldsymbol{x} \in \mathbb{R}^n$$

- Search direction $\boldsymbol{d} = -\frac{\nabla f(\boldsymbol{x})}{\| \nabla f(\boldsymbol{x}) \|}$, which is the steepest descent direction at the point $\boldsymbol{x}$
- Stepsize λ can be found with line search
- A new point is of the form $\boldsymbol{x}_+ = \boldsymbol{x} + \lambda \boldsymbol{d}$, where $f(\boldsymbol{x}_+) < f(\boldsymbol{x})$ until the optimum is reached

About MGDA

- Assume that in problem (1) we have
  - All objective functions f_i are continiously differentiable
  - Problem has no constrains and feasible set is $\mathbb{R}^n$
- The main idea is to find a common descent direction for all objective functions by defining the convex hull of gradients of objective functions and finding the minimum norm element of the convex hull
- With this method we obtain one Pareto stationary solution

Pareto stationary

Definition

Objective functions f_i are said to be Pareto stationary at the point $\boldsymbol{x}^*$ if

$$\sum_{i=1}^{k} \alpha_i \nabla f_i(\boldsymbol{x}^*) = 0, \quad \alpha_i \geq 0 \quad \forall i, \quad \sum_{i=1}^{k} \alpha_i = 1.$$

- Pareto stationarity is a necessary condition for Pareto optimality

Theorem

Let objective funtions f_i, $i = 1, \dots, k$ be continiously differentiable and

$$U = \{\boldsymbol{w} \in \mathbb{R}^n \mid \boldsymbol{w} = \sum_{i=1}^{k} \alpha_i \nabla f_i(\boldsymbol{x}), \alpha_i \geq 0 \; \forall \, i, \sum_{i=1}^{k} \alpha_i = 1\}$$

Let vector $\boldsymbol{w}^*$ be a minimum norm element of convex hull of U i.e. $\boldsymbol{w}^* = \text{argmin}\{\| \boldsymbol{w} \| \mid \boldsymbol{w} \in \text{conv } U\}$. Then

1. *either $\boldsymbol{w}^* = 0$ and objective functions f_i, $i = 1, \dots, k$ are Pareto stationary at the point $\boldsymbol{x}^*$*
2. *or $\boldsymbol{w}^* \neq 0$ and $-\boldsymbol{w}^*$ is common descent direction for all objective functions. Additionally, if $\boldsymbol{w}^* \in U$ then $\boldsymbol{u}^T \boldsymbol{w}^* = \| \boldsymbol{w}^* \|^2$ for all $\boldsymbol{u} \in \text{conv } U$.*

Search direction

- To find the vector $\boldsymbol{w}$ we need to solve the problem

$$\min \quad \|\sum_{i=1}^{k} \alpha_i \nabla f_i(\boldsymbol{x})\|^2$$

$$\text{s. t.} \quad \alpha_i \geq 0 \, \forall \, i, \quad \sum_{i=1}^{k} \alpha_i = 1.$$

- In case of two objective functions we have

$$\alpha_1 = \begin{cases} \frac{\|\boldsymbol{v}\|^2 - \boldsymbol{u}^T\boldsymbol{v}}{\|\boldsymbol{u}\|^2 + \|\boldsymbol{v}\|^2 - 2\boldsymbol{u}^T\boldsymbol{v}}, & \text{if } \boldsymbol{u}^T\boldsymbol{v} < \min\{\|\boldsymbol{u}\|^2, \|\boldsymbol{v}\|^2\} \\ 0, & \text{if } \min\{\|\boldsymbol{u}\|, \|\boldsymbol{v}\|\} = \|\boldsymbol{v}\| \\ 1, & \text{if } \min\{\|\boldsymbol{u}\|, \|\boldsymbol{v}\|\} = \|\boldsymbol{u}\| \end{cases}$$

where $\boldsymbol{u} = \nabla f_1(\boldsymbol{x})$, $\boldsymbol{v} = \nabla f_2(\boldsymbol{x})$ and $\alpha_2 = 1 - \alpha_1$.

Stepsize

- Stepsize λ is the largest strictly positive real number such that all the functions $g_i(t) = f_i(\boldsymbol{x} - t\boldsymbol{w})$ are monotonically decreasing over the interval $[0, t]$.
- A new point $\boldsymbol{x}_+$ is $\boldsymbol{x}_+ = \boldsymbol{x} - \lambda\boldsymbol{w}$
- Algorithm stops when $\boldsymbol{w} = 0$

About MGDA for non-smooth convex problems

- Assume that in the problem (1)
  - Objective functions f_i are convex functions, not necessarily differentiable
  - Problem has no constrains and the feasible set is $\mathbb{R}^n$
  - The subdifferentials of objective functions f_i are known
- The basic idea is the same as before but now instead of gradients we use the steepest descent subgradients

Subdifferential and subgradient

Definition

Let function f be convex. A subdifferential of f at the point $\boldsymbol{x}$ is a set

$$\partial f(\boldsymbol{x}) = \{\xi \in \mathbb{R}^n \mid f'(\boldsymbol{x}; \boldsymbol{d}) \geq \xi^T \boldsymbol{d}\}$$

for all $\boldsymbol{d} \in \mathbb{R}^n$. A vector $\xi \in \partial f(\boldsymbol{x})$ is called subgradient of function f at the point $\boldsymbol{x}$.

Theorem

Let function $f : \mathbb{R}^n \longrightarrow \mathbb{R}$ be convex and differentiable at the point $\boldsymbol{x} \in \mathbb{R}^n$. Then

$$\partial f(\boldsymbol{x}) = \{\nabla f(\boldsymbol{x})\}.$$

Steepest descent direction

- To determine the steepest descent direction we need to solve the problem

$$\min \quad f'(\boldsymbol{x};\boldsymbol{d}) \qquad \text{s. t.} \quad \|\boldsymbol{d}\| = 1 \tag{2}$$

- The problem (2) can be written

$$\min \quad \|\xi\| \qquad \text{s. t.} \quad \xi \in \partial f(\boldsymbol{x})$$

- Thus the steepest descent direction at the point $\boldsymbol{x}$ is the smallest norm subgradient in the subdifferential $\partial f(\boldsymbol{x})$

Algorithm

- To obtain a search direction we need to define the steepest descent subgradient for every objective function at the point $\boldsymbol{x}$. After that we can solve problem

$$\begin{aligned} \min \quad & \| \sum_{i=1}^{k} \alpha_i \xi_i \|^2 \\ \text{s. t.} \quad & \alpha_i \geq 0 \, \forall \, i, \quad \sum_{i=1}^{k} \alpha_i = 1 \end{aligned}$$

where $\xi_i \in \partial f_i(\boldsymbol{x})$ is the steepest descent subgradient.
- Stepsize and new point are determined as in smooth case

- With this method we obtain one Pareto stationary solution since

$$\sum_{i=1}^{k} \alpha_i \xi_i = 0, \quad \alpha_i \geq 0 \quad \text{and} \quad \sum_{i=1}^{k} \alpha_i = 1.$$

- In convex case Pareto stationarity is necessary and sufficient condition to Pareto optimality

  $\Rightarrow$ We obtain one Pareto optimal solution

Example

Consider the following problem:

$$\begin{aligned} \min \quad & \{f_1(\boldsymbol{x}), f_2(\boldsymbol{x})\} \\ \text{s. t.} \quad & \boldsymbol{x} \in \mathbb{R}^2 \end{aligned}$$

where

$$\begin{aligned} f_1(\boldsymbol{x}) &= \max\{x_1^2 + (x_2 - 1)^2,\ (x_1 + 1)^2\} \\ f_2(\boldsymbol{x}) &= \max\{2x_1 - x_2, x_1^2 + x_2\} \end{aligned}$$

and starting point $\boldsymbol{x}_0 = (1, 0.5)$.

The steepest descent subgradients at the point $\boldsymbol{x}_0$:

- $\xi_1 = (4, 0)$
- To determine ξ_2 we need to find $\rho \in [0, 1]$ such that we find the minimum norm of subdifferential

$$\partial f_2(\boldsymbol{x}_0) = \rho(2, -1) + (1 - \rho)(2, 1) = (2, 1 - 2\rho).$$

To obtain the steepest descent subgradient $\rho = \frac{1}{2}$ and $\xi_2 = (2, 0)$.

A common descent search direction:

- $\boldsymbol{w} = \alpha\xi_1 + (1-\alpha)\xi_2$
- We obtain α from formula

$$\alpha = \begin{cases} \frac{\|\xi_2\|^2 - \xi_1^T\xi_2}{\|\xi_1\|^2 + \|\xi_2\|^2 - 2\xi_1^T\xi_2}, & \text{if } \xi_1^T\xi_2 < \min\{\|\xi_1\|^2, \|\xi_2\|^2\} \\ 0, & \text{if } \min\{\|\xi_1\|, \|\xi_2\|\} = \|\xi_2\| \\ 1, & \text{if } \min\{\|\xi_1\|, \|\xi_2\|\} = \|\xi_1\| \end{cases}$$

- Thus $\boldsymbol{w} = (2, 0)$

Stepsize:

- Function $g_1(t) = f_1(\boldsymbol{x}_0 - t\boldsymbol{w})$ is descent if $t \in [0, 0.6875]$
- Function $g_2(t) = f_2(\boldsymbol{x}_0 - t\boldsymbol{w})$ is descent if $t \in [0, 0.5]$

  $\Rightarrow$ Stepsize $\lambda = 0.5$

New point:

- $\boldsymbol{x}_1 = \boldsymbol{x}_0 - \lambda\boldsymbol{w} = (0, 0.5)$

k	ξ_1	ξ_2	α	$\boldsymbol{w}$	λ	$\boldsymbol{x}_{k+1}$
0	$(4, 0)$	$(2, 0)$	0	$(2, 0)$	0.5	$(0, 0.5)$
1	$(2, 0)$	$(0, 1)$	0.2	$(0.4, 0.8)$	0.403	$(-0.161, 0.178)$
2	$(1.678, 0)$	$(-0.322, 1)$	0.329	$(0.336; 0.671)$		

Solution: $\boldsymbol{x}^* = (-0.161, 0.178)$, $f_1(\boldsymbol{x}^*) \approx 0.704$ and $f_2(\boldsymbol{x}^*) \approx 0.203$

Future work

- The biggest drawback is that we need to know whole subdifferential
- The aim is to find an effective way to get a common descent direction for all objective functions
- We do not need more than one arbitrary subgradient of function

Thank you for your attention!