Decision Methods and Models

Master's Degree in Computer Science

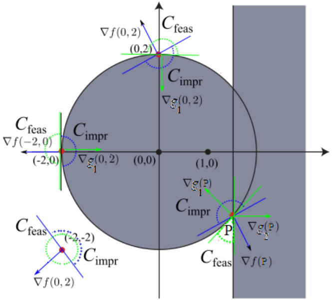
Roberto Cordone

DI - Università degli Studi di Milano

Schedule:	Thursday 16.30 - 18.30 in Aula Magna (CS department)
	Friday 12.30 - 14.30 in classroom 301
Office hours:	on appointment
E-mail:	roberto.cordone@unimi.it
Web page:	https://homes.di.unimi.it/cordone/courses/2024-mmd/2024-mmd.html
Ariel site:	https://myariel.unimi.it/course/view.php?id=4467

First geometric interpretation

$$
\begin{aligned}
\min f(x) &= (x_1 - 1)^2 + x_2^2 \\
g_1(x) &= -x_1^2 - x_2^2 + 4 \leq 0 \\
g_2(x) &= x_1 - 3/2 \leq 0
\end{aligned}
$$



The objective f worsens when moving farther from $(1,0)$

- Point $(0,2)$: g_1 active (move up), g_2 nonactive (free):
  $C_{\text{feas}}(x)$ and $C_{\text{impr}}(x)$ intersect
- Point $(-3/2, -\sqrt{7}/2)$: g_1 active (move down-right), g_2 active (move left):
  $C_{\text{feas}}(x)$ and $C_{\text{impr}}(x)$ do not intersect: candidate (globally optimal)
- Point $(-2,-2)$: g_1 and g_2 nonactive (free):
  $C_{\text{feas}}(x)$ and $C_{\text{impr}}(x)$ do not intersect
- Point $(-2,0)$: g_1 active (move left), g_2 nonactive (free):
  $C_{\text{feas}}(x)$ and $C_{\text{impr}}(x)$ (open!) do not intersect:
  candidate (but not even locally optimal)

A filtering approach

We have modified the filtering approach using analytic conditions

- nonregular points are identified and considered as candidates
- in regular points, analytic conditions identify feasible arcs
- analytic conditions identify improving arcs

$X^{KKT} := X \setminus$ NonRegular(g, X);
For each $x \in X^{KKT}$ *(continuous set for x)*
  For each $p \in \mathbb{R}^n : \nabla g_j(x)^T p \leq 0, \forall j \in J_a(x)$ *(arc $\xi(\alpha)$ in x feasible for X)*
    If $\nabla f(\tilde{x})^T p_\xi < 0$ *($\xi(\alpha)$ is improving in x for $f(\cdot)$)*
      then $X^{KKT} := X^{KKT} \setminus \{x\}$
$X^{KKT} := X^{KKT} \cup$ NonRegular(g, X);
Return X^{KKT}

But still we need to loop on continuous sets

- feasible regular solutions $x \in X^{KKT}$
- potential tangent vectors $p \in \mathbb{R}^n$

We need a powerful change of perspective

Farkas' lemma

Given $f \in \mathbb{R}^n$ and $g_j \in \mathbb{R}^n$ with $j = 1, \ldots, m$

$$\exists \mu_j \geq 0 : f = \sum_{j=1}^{m} \mu_j g_j \;\Leftrightarrow\; p^T f \leq 0 \;\forall p : p^T g_j \leq 0 \;\forall j$$

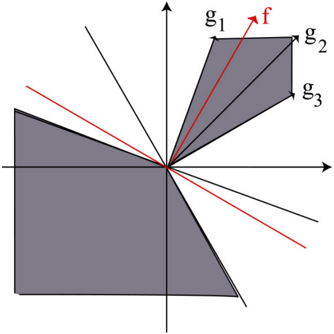
Vector f is conic combination of vectors g_j if and only if all vectors pointing away from all g_j point away from f

The direct implication is trivial

The converse implication is hard to prove (*we give it for granted*)

We apply the converse implication to

- $f = -\nabla f(x^*)$ (antigradient)
- $g_j = \nabla g_j(x^*)$



Farkas' lemma and local optimality

A vector p with $p^T \nabla g_j(x) \leq 0, \; \forall j \in J_a(x)$ is tangent to a feasible arc

We remove point x when $\nabla f(x)^T p < 0$ for all such vectors

Therefore, x is candidate when $\nabla f(x)^T p \geq 0$ for all feasible tangents

$$p^T \nabla f(x^*) \leq 0 \text{ for all } p : p^T \nabla g_j(x^*) \leq 0, \; \forall j \in J_a(x)$$

that is exactly the second expression mentioned in Farkas' lemma

Replace it with the equivalent first expression

A regular point x is candidate for local optimality when

$$\exists \mu_j \geq 0 : \nabla f(x) + \sum_{j \in J_a(x)} \mu_j \nabla g_j(x) = 0$$

Don't loop on all x and p: solve a system of equations in μ and x

Second geometric interpretation

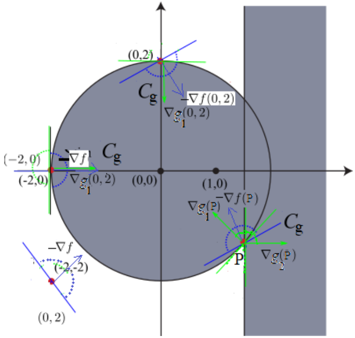
- The antigradient $-\nabla f(x)$ is the direction of quickest improvement
- The gradients of the active constraint $\nabla g_j(x)$ are the directions of quickest violation
- Multipliers $\mu_j \geq 0$ let combination $\sum_{j \in J_a(x)} \mu_j \nabla g_j(x)$ describe a cone, denoted as C_g, the cone of the gradients of the active constraints
- In candidate points, the antigradient of the objective falls in C_g

$$-\nabla f(x) = \sum_{j \in J_a(x)} \mu_j \nabla g_j(x)$$

When $-\nabla f(x) \in C_g(x)$, the objective improves only violating constraints

Hence, no vector is tangent to a feasible and improving arc

Example



The antigradient $-\nabla f$ points towards $(1, 0)$

- Point $(0, 2)$: g_1 active; $C_g(x)$ is the half-line pointing downwards $-\nabla f(x)$ does not belong to $C_g(x)$
- Point $(-3/2, -\sqrt{7}/2)$: g_1 and g_2 active; C_g points up, right, partly left: $-\nabla f(x)$ falls inside $C_g(x)$: candidate (globally optimal)
- Point $(-2, -2)$: g_1 and g_2 nonactive; $C_g(x)$ is empty: $-\nabla f(x)$ does not belong to $C_g(x)$
- Point $(-2, 0)$: g_1 active, g_2 nonactive: $C_g(x)$ is the half-line pointing right $-\nabla f(x)$ falls inside $C_g(x)$: candidate (but not even locally optimal)

Minor tweaks (1)

The analytic conditions still have a problem:

$$\nabla f(x) + \sum_{j \in J_a(x)} \mu_j \nabla g_j(x) = 0$$

The sum applies only to the constraints active in an unknown point x!

How to find them?

The conditions are reformulated

1. adding also the gradients of the nonactive constraints

$$\nabla f(x) + \sum_{j=1}^{m} \mu_j \nabla g_j(x) = 0$$

2. forcing to zero the multipliers of the nonactive constraints

$$\mu_j g_j(x) = 0, \text{ for } j = 1, \ldots, m$$

These equalities are called the complementarity conditions, as they are equivalent to $g_j(x) < 0 \Rightarrow \mu_j = 0$

Minor tweaks (2)

Equality constraints can be replaced by pairs of inequalities

$$h_i(x) = 0 \Leftrightarrow \begin{cases} g_{j_i'}(x) = h_i(x) & \leq 0 \\ g_{j_i''}(x) = -h_i(x) & \leq 0 \end{cases}$$

but it is simpler to

1. use a single multiplier λ_i
2. relax the nonnegativity condition on the multiplier: $\lambda_i \in \mathbb{R}$

In fact

$$\begin{cases} \nabla g_{j_i'}(x) & = \nabla h_i(x) \\ \nabla g_{j_i''}(x) & = -\nabla h_i(x) \end{cases}$$

and

$$\mu_{j_i'} \nabla g_{j_i'}(x) + \mu_{j_i''} \nabla g_{j_i''}(x) = \left(\mu_{j_i'} - \mu_{j_i''}\right) \nabla h_i(x) = \lambda_i \nabla h_i(x)$$

with $\mu_{j_i'} \geq 0$, $\mu_{j_i''} \geq 0$ and $\mu_{j_i'} - \mu_{j_i''} = \lambda_i$ free in $\mathbb{R}$

Karush-Kuhn-Tucker conditions

Theorem

Let

- $f \in C^1(X)$
- $X = \{x \in \mathbb{R}^n : h_i(x) = 0 \text{ with } h_i \in C^1(X) \text{ for } i = 1, \dots, s,$
  $g_j(x) \leq 0, \text{ with } g_j \in C^1(X) \text{ for } j = 1, \dots, m\}$
- x^* is a locally optimal point for f in X
- x^* is a regular point in X

Then there exist multipliers $\lambda_i \in \mathbb{R}$ and $\mu_j \geq 0$ such that

$$\nabla f(x^*) + \sum_{i=1}^{s} \lambda_i \nabla h_i(x^*) + \sum_{j=1}^{m} \mu_j \nabla g_j(x^*) = 0$$

$$\begin{aligned} h_i(x^*) &= 0 & i &= 1, \dots, s \\ \mu_j g_j(x^*) &= 0 & j &= 1, \dots, m \\ g_j(x^*) &\leq 0 & j &= 1, \dots, m \\ \mu_j &\geq 0 & j &= 1, \dots, m \end{aligned}$$

It is no longer required to scan continuous sets of points, arcs or vectors: we just need to solve a system of equalities and inequalities

Balance between equations and variables

The system consists of $n + s + m$ equations

- n equations from Farkas' lemma
- s equations from equality constraints
- m equations from the complementarity conditions

imposed on $n + s + m$ variables

- n variables to determine the solution x
- s variables to determine the multiplier vector λ
- m variables to determine the multiplier vector μ

and is therefore balanced

The $2m$ inequalities

- m constraints on the solution
- m nonnegativity conditions on the multipliers

remove solutions, but do not decrease the freedom degrees

The number of solutions is probably finite

Relevant particular cases

- unconstrained problems are solved by the classical condition on first derivatives

$$\nabla f(x) = 0 \Leftrightarrow \frac{\partial f}{\partial x_i} = 0 \text{ for all } i = 1, \ldots, n$$

- linear programming is solved imposing:
  - Farkas' lemma, that corresponds to the dual constraints
  - the complementarity conditions, that correspond to complementary slackness
  - the feasibility constraints, that correspond to the primal constraints
  - the nonnegativity conditions, that correspond to the nonnegativity of the dual variables
- on discrete problems the conditions are correct, but useless: the integrality constraint

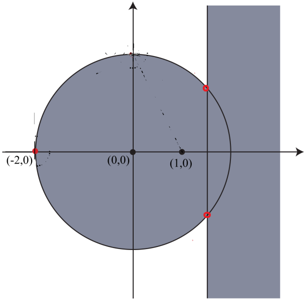
$$X \subseteq \mathbb{Z}^n \Leftrightarrow h_i(x) = \sin(\pi x_i) = 0$$

occurs explicitly in the system and every feasible point is candidate: the KKT conditions reduce to the exhaustive algorithm

This is obvious: isolated points are locally optimal!

Example

$$\begin{aligned} \min f(x) &= (x_1 - 1)^2 + x_2^2 \\ g_1(x) &= -x_1^2 - x_2^2 + 4 \le 0 \\ g_2(x) &= x_1 - 3/2 \le 0 \end{aligned}$$



See the detailed computations in the lecture notes