Modern Optimization Techniques

4. Inequality Constrained Optimization / 4.1. Primal Methods

Lars Schmidt-Thieme

Information Systems and Machine Learning Lab (ISMLL)
Institute for Computer Science
University of Hildesheim, Germany

Syllabus

Mon. 28.10.	(0)	0. Overview
		1. Theory
Mon. 4.11.	(1)	1. Convex Sets and Functions
		2. Unconstrained Optimization
Mon. 11.11.	(2)	2.1 Gradient Descent
Mon. 18.11.	(3)	2.2 Stochastic Gradient Descent
Mon. 25.11.	(4)	2.3 Newton's Method
Mon. 2.12.	(5)	2.4 Quasi-Newton Methods
Mon. 19.12.	(6)	2.5 Subgradient Methods
Mon. 16.12.	(7)	2.6 Coordinate Descent
	—	— *Christmas Break* —
		3. Equality Constrained Optimization
Mon. 6.1.	(8)	3.1 Duality
Mon. 13.1.	(9)	3.2 Methods
		4. Inequality Constrained Optimization
Mon. 20.1.	(10)	4.1 Primal Methods
Mon. 27.1.	(11)	4.2 Barrier and Penalty Methods
Mon. 3.2.	(12)	4.3 Cutting Plane Methods

Outline

1. Inequality Constrained Minimization Problems

2. Maintaining Strict Inequality Constraints

3. Gradient Projection Method for Affine Equality Constraints

4. Active Set Methods: General Strategy

5. Gradient Projection Method for Affine Inequality Constraints

Outline

Inequality Constrained Minimization Problem

A problem of the form:

$$\begin{aligned} \arg\min_{x \in \mathbb{R}^N} \; & f(\mathbf{x}) \\ \text{subject to} \; & g_p(\mathbf{x}) = 0, \quad p = 1, \ldots, P \\ & h_q(\mathbf{x}) \leq 0, \quad q = 1, \ldots, Q \end{aligned}$$

where:

- $f : \mathbb{R}^N \to \mathbb{R}$ **objective function**
- $g_1, \ldots, g_P : \mathbb{R}^N \to \mathbb{R}$ **equality constraints**
- $h_1, \ldots, h_Q : \mathbb{R}^N \to \mathbb{R}$ **inequality constraints**
- A feasible optimal $\mathbf{x}^*$ exists, $p^* := f(\mathbf{x}^*)$

Inequality Constrained Minimization Problem / Convex

A problem of the form:

$$\begin{aligned} \arg\min_{x \in \mathbb{R}^N} \quad & f(\mathbf{x}) \\ \text{subject to} \quad & A\mathbf{x} - a = 0 \\ & h_q(\mathbf{x}) \leq 0, \quad q = 1, \ldots, Q \end{aligned}$$

where:

- $f : \mathbb{R}^N \rightarrow \mathbb{R}$ **convex** and **twice differentiable**
- $A \in \mathbb{R}^{P \times N}, a \in \mathbb{R}^P$: P **affine** equality constraints
- $h_1, \ldots, h_Q : \mathbb{R}^N \rightarrow \mathbb{R}$ **convex** and **twice differentiable**
- A feasible optimal $\mathbf{x}^*$ exists, $p^* := f(\mathbf{x}^*)$

Inequality Constrained Minimization Problem / Affine

$$\begin{aligned} \arg\min_{x \in \mathbb{R}^N} \; & f(\mathbf{x}) \\ \text{subject to} \; & A\mathbf{x} - a = 0 \\ & B\mathbf{x} - b \leq 0 \end{aligned}$$

where:

- $f : \mathbb{R}^N \to \mathbb{R}$ convex and twice differentiable
- $A \in \mathbb{R}^{P \times N}, a \in \mathbb{R}^P$: P affine equality constraints
- $B \in \mathbb{R}^{Q \times N}, b \in \mathbb{R}^Q$: Q **affine** inequality constraints
- A feasible optimal $\mathbf{x}^*$ exists, $p^* := f(\mathbf{x}^*)$

Primal Methods

- Primal methods tackle the problem directly,
  - starting from a feasible point $x^{(0)}$
  - staying all time within the feasible area
    - i.e., all $x^{(k)}$ are feasible

Advantages:

1. If stopped early,
   yields a feasible point with often already small objective value.
2. If converged,
   also for non-convex objectives yields at least a local optimum.
3. Generally applicable, as they do not rely on special problem structure.

Active Set Methods / General Idea

- split inequality constraints into
  - active constraints: $h_q(x) = 0$
  - inactive constraints: $h_q(x) < 0$

- enhance methods for equality constraints to
  - retain strict inequality constraints $h_q(x) < 0$
    - by taking small steps
  - to stop, once they hit an inequality constraint $h_q(x) = 0$

Further procedure:

1. enhance backtracking to respect strict inequality constraints
2. enhance gradient projection to respect strict inequality constraints
   - gradient descent with affine equality constraints
3. sketch the general strategy of active set methods

Outline

Backtracking Line Search (Review)

```
linesearch-bt(f, ∇f, x, Δx; α, β):
  μ := 1
  Δf := α∇f(x)^T Δx
  while f(x + μΔx) > f(x) + μΔf:
    μ := βμ
  return μ
```

where

- $f : \mathbb{R}^N \rightarrow R, \nabla f : \mathbb{R}^N \rightarrow \mathbb{R}$: objective function and its gradient
- $x \in \mathbb{R}^N$: current point
- $\Delta x \in \mathbb{R}^N$: update/search direction
- $\alpha \in (0, 0.5)$: minimum descent steepness
- $\beta \in (0, 1)$: stepsize shrinkage factor

Backtracking Line Search / Inequality Constraints

linesearch-bt-ineq$(f, \nabla f, h, x, \Delta x; \alpha, \beta)$:
  $\mu := 1$
  $\Delta f := \alpha \nabla f(\mathbf{x})^T \Delta x$
  while $f(x + \mu \Delta x) > f(x) + \mu \Delta f$ or not $h(x + \mu \Delta x) \leq 0$:
    $\mu := \beta \mu$
  return μ

where

- $f : \mathbb{R}^N \rightarrow R, \nabla f : \mathbb{R}^N \rightarrow \mathbb{R}$: objective function and its gradient
- $x \in \mathbb{R}^N$: current point, feasible: $h(x) \leq 0$
- $\Delta x \in \mathbb{R}^N$: update/search direction
- $\alpha \in (0, 0.5)$: minimum descent steepness
- $\beta \in (0, 1)$: stepsize shrinkage factor
- $h : \mathbb{R}^N \rightarrow \mathbb{R}^Q$: Q inequality constraints: $h(x) \leq 0$

Backtracking Line Search / Affine Inequality Constraints

For affine inequality constraints

$$h(x) = Bx - b \leq 0$$

feasibility of an update can be guaranteed by a maximal stepsize:

$$\begin{aligned}
h(x + \mu\Delta x) = \\
B(x + \mu\Delta x) - b \leq 0 \\
\mu B \Delta x &\leq -(Bx - b) \\
\mu (B\Delta x)_q &\leq -(Bx - b)_q \quad \forall q \in \{1, \ldots, Q\} \\
\mu &\leq \frac{-(Bx-b)_q}{(B\Delta x)_q} \quad \forall q \in \{1, \ldots, Q\} : (B\Delta x)_q > 0 \\
\mu &\leq \min\{\frac{-(Bx-b)_q}{(B\Delta x)_q} \mid q \in \{1, \ldots, Q\} : (B\Delta x)_q > 0\} \\
&=: \mu_{\max}
\end{aligned}$$

Backtracking Line Search / Affine Inequality Constraints

linesearch-bt-affineq$(f, \nabla f, B, b, x, \Delta x; \alpha, \beta)$:

$\quad \mu := \min\{\frac{-(Bx-b)_q}{(B\Delta x)_q} \mid q \in \{1, \ldots, Q\} : (B\Delta x)_q > 0\}$

$\quad \Delta f := \alpha \nabla f(\mathbf{x})^T \Delta x$

$\quad$ while $f(x + \mu \Delta x) > f(x) + \mu \Delta f$:

$\quad\quad \mu := \beta \mu$

$\quad$ return μ

where

- $f : \mathbb{R}^N \to R, \nabla f : \mathbb{R}^N \to \mathbb{R}$: objective function and its gradient
- $x \in \mathbb{R}^N$: current point, feasible: $Bx - b \leq 0$
- $\Delta x \in \mathbb{R}^N$: update/search direction
- $\alpha \in (0, 0.5)$: minimum descent steepness
- $\beta \in (0, 1)$: stepsize shrinkage factor
- $B \in \mathbb{R}^{Q \times N}, b \in \mathbb{R}^Q$: Q affine inequality constraints: $Bx - b \leq 0$

Outline

Right Inverse Matrix

For $A \in \mathbb{R}^{N \times M}$ $(N \leq M)$ with full rank,
the right inverse of A is

$$A^{-1}_{\text{right}} = A^T (AA^T)^{-1}$$

Proof:

$$AA^{-1}_{\text{right}} = AA^T (AA^T)^{-1} = I$$

Nullspace Projection

For $A \in \mathbb{R}^{N \times M}$ $(N \leq M)$ with full rank, the matrix

$$F := I - A^{-1}_{\text{right}} A = I - A^T (AA^T)^{-1} A$$

is a projection onto the nullspace of A:

$$\{x \in \mathbb{R}^M \mid Ax = 0\} = \{Fx' \mid x' \in \mathbb{R}^M\}$$

Proof:

" $\supseteq$ " : $\quad AFx' = A(I - A^{-1}_{\text{right}} A)x' \quad = (A - A)x' = 0$

" $\subseteq$ " : show: for any x with $Ax = 0$, there exists $x' : x = Fx'$

$$x' := x : Fx' = Fx = (I - A^T (AA^T)^{-1} A)x \quad = x - A^T (AA^T)^{-1} Ax$$
$$= x - 0 = x$$

Gradient Projection Method / Affine Equality Constraints

```
min-gp-affeq(f, ∇f, A, x^(0), μ, ε, K):
    F := I − A^T (A A^T)^{-1} A
    for k := 1, ..., K:
        Δx^(k−1) := −F^T ∇f(x^(k−1))
        if ||Δx^(k−1)|| < ε:
            return x^(k−1)
        μ^(k−1) := μ(f, x^(k−1), Δx^(k−1))
        x^(k) := x^(k−1) + μ^(k−1) Δx^(k−1)
    return "not converged"
```

where

- $A \in \mathbb{R}^{P \times N}, a \in \mathbb{R}^P$: P affine equality constraints
- $x^{(0)}$ feasible starting point, i.e., $Ax^{(0)} - a = 0$

Grad. Proj. Meth. / Aff. Eq. Cstr. + strict In.eq. Constr

```
min-gp-affeq-strictineq(f, ∇f, A, h, x^(0), μ, ε, K):
  F := I − A^T (A A^T)^{-1} A
  for k := 1, ..., K:
    Δx^(k−1) := −F^T ∇f(x^(k−1))
    if ||Δx^(k−1)|| < ε:
      return x^(k−1)
    μ^(k−1) := μ(f, h, x^(k−1), Δx^(k−1))
    x^(k) := x^(k−1) + μ^(k−1) Δx^(k−1)
    if ∃q ∈ {1, ..., Q} : h_q(x^(k)) = 0 :
      return x^(k)
  return "not converged"
```

where

- $A \in \mathbb{R}^{P \times N}, a \in \mathbb{R}^P$: P affine equality constraints
- $x^{(0)}$ strictly feasible starting point, i.e., $h(x^{(0)}) < 0$, $Ax^{(0)} - a = 0$
- $\mu(\ldots, h, \ldots)$ stepsize controller that retains inequality constraints h
- $h : \mathbb{R}^N \rightarrow \mathbb{R}^Q$: Q inequality constraints: $h(x) \leq 0$

Outline

Active Set Method / Idea

- split inequality constraints into
  - active constraints: $h_q(x) = 0$
  - inactive constraints: $h_q(x) < 0$
- minimize on the feasible subspace retaining the active constraints
  - add active inequality constraints (temporarily) to the equality constraints: $\tilde{g}$
  - make small steps μ s.t. inactive constraints remain inactive
    - stop if a step hits one of the inactive constraints, activating them.
- once the minimum on the subspace of the current active constraints is found,
  - if we had to stop because of hitting an active constraint:
    - add one of the hit constraints to the active constraints
  - otherwise:
    - inactivate one of the active constraints
      one on whos interior side the objective is decreasing ($\lambda_q < 0$)

Active Set Methods / General Strategy

min-activeset$(f, g, h, x^{(0)}, K, \text{min-eq})$:
  $\mathcal{Q} := \{q \in \{1, \ldots, Q\} \mid h_q(x^{(0)}) = 0\}$
  $\tilde{g} := \begin{pmatrix} g \\ h_{\mathcal{Q}} \end{pmatrix}, \quad \tilde{h} := h_{\{1,\ldots,Q\} \setminus \mathcal{Q}}$
  for $k := 1, \ldots, K$:
    $x^{(k)} := \text{min-eq}(f, \tilde{g}, \tilde{h}, x^{(k-1)})$
    if $\exists q \in \{1, \ldots, Q\} \setminus \mathcal{Q} : h_q(x) = 0$:
      $\mathcal{Q} := \mathcal{Q} \cup \{q\}$ for an arbitrary $q \in \{1, \ldots, Q\} \setminus \mathcal{Q}$ with $h_q(x) = 0$
      $\tilde{g} := \begin{pmatrix} g \\ h_{\mathcal{Q}} \end{pmatrix}, \quad \tilde{h} := h_{\{1,\ldots,Q\} \setminus \mathcal{Q}}$
    else :
      if $|\mathcal{Q}| = 0$:
        return $x^{(k)}$
      compute Lagrange multipliers λ_q for $h_q, q \in \mathcal{Q}$
      if $\lambda \geq 0$:
        return $x^{(k)}$
      $\mathcal{Q} := \mathcal{Q} \setminus \{q\}$ for an arbitrary $q \in \mathcal{Q}$ with $\lambda_q < 0$
      $\tilde{g} := \begin{pmatrix} g \\ h_{\mathcal{Q}} \end{pmatrix}, \quad \tilde{h} := h_{\{1,\ldots,Q\} \setminus \mathcal{Q}}$
  return "not converged"

where

- $g : \mathbb{R}^N \rightarrow \mathbb{R}^P$: P equality constraints: $g(x) = 0$
- $h : \mathbb{R}^N \rightarrow \mathbb{R}^Q$: Q inequality constraints: $h(x) \leq 0$
- $x^{(0)}$ feasible starting point, i.e., $g(x) = 0, h(x) \leq 0$
- min-eq: solver for equality constraints and strict inequality constraints, e.g., min-qp-affeq-strictineq

Computing the Lagrange Multipliers (line 12)

complementary slackness:

$$\lambda_q h_q(x) = 0 \quad \rightsquigarrow \lambda_q = 0 \ \forall q \notin \mathcal{Q}$$

stationarity:

$$\nabla f(\mathbf{x}) + \sum_{p=1}^{P} \nu_p \nabla g_p(\mathbf{x}) + \sum_{q=1}^{Q} \lambda_q \nabla h_q(\mathbf{x}) = \nabla f(\mathbf{x}) + \sum_{p=1}^{\tilde{P}} \tilde{\nu}_p \nabla \tilde{g}_p(\mathbf{x}) = 0$$

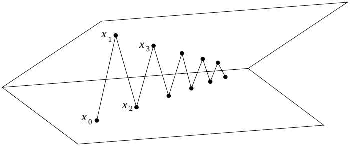
$\rightsquigarrow$ solve LSE

$$(\nabla \tilde{g}_1(\mathbf{x}), \ldots, \nabla \tilde{g}_{\tilde{P}}(\mathbf{x})) \begin{pmatrix} \tilde{\nu}_1 \\ \vdots \\ \tilde{\nu}_{\tilde{P}} \end{pmatrix} = -\nabla f(x)$$

$$\lambda_q := \tilde{\nu}_p \quad \text{for } p \in \{1, \ldots, \tilde{P}\} : \tilde{g}_p = h_q, \quad q \in \mathcal{Q}$$

Active Set Method / Remarks

- Limitation: To work with non-linear inequality constraints h_q, the active set method requires an equality-constrained optimizer min-eq that can cope with **non-linear equality constraints**.
  - because active inequality constraints h_q are used as equality constraints $\tilde{g}_p$.
- The active set method can be accelerated by solving the equality constrained problem only approximately: ϵ
  - but for the risk of zigzagging



[Griva et al., 2009, p.570]

Convergence

Theorem (Active Set Theorem)

If for every subset $\mathcal{Q}$ of inequality constraints the problem

$$\begin{aligned} \arg\min_{x \in \mathbb{R}^N} \; & f(x) \\ \textit{subject to } Ax - a &= 0 \\ B_{\mathcal{Q}} x - b_{\mathcal{Q}} &= 0 \\ B_{\bar{\mathcal{Q}}} x - b_{\bar{\mathcal{Q}}} &< 0, \quad \bar{\mathcal{Q}} := \{1, \ldots, Q\} \setminus \mathcal{Q} \end{aligned}$$

is well-defined with a unique nondegenerate solution (i.e.,$\lambda_q \neq 0 \; \forall q \in \mathcal{Q}$), then the active set method converges to the solution of the inequality constrained problem.

Proof:

- After the minimum over the subspace defined by an active set has been found,
- the function value further decreases when removing a constraint.
- Thus the algorithm cannot possibly return to the same active set.
- As there are only finite many possible active sets, it eventually will terminate.

Outline

Gradient Projection / Idea

- Gradient Projection:
  - use the active set strategy for Gradient Descent (to solve the equality constrained subproblems)

- putting everything together
  - esp. for affine constraints

Gradient Projection / Idea

- split inequality constraints into
  - active constraints: $(Bx - b)_q = 0$
  - inactive constraints: $(Bx - b)_q < 0$
- find an update direction Δx that retains this state of the inequality constraints
  - add active inequality constraints (temporarily) to the equality constraints: $\tilde{A}, \tilde{a}$
  - make small steps μ s.t. inactive constraints remain inactive:

$$(B(x + \mu\Delta x) - b)_q \leq 0 \rightsquigarrow \mu \leq \frac{-(Bx - b)_q}{(B\Delta x)_q}, \quad \text{for } (B\Delta x)_q > 0$$

    - $x + \mu\Delta x$ may hit one of the inactive constraints, activating them.
- once the minimum on the subspace of the current active constraints is found,
  - inactivate one of the active constraints
    - one on whos interior side the objective is decreasing ($\lambda_q < 0$)

Gradient Projection / Affine Constraints

1 **min-gp-aff**($f, A, a, B, b, x^{(0)}, \mu, \epsilon, K$):
2 $\quad \mathcal{Q} := \{q \in \{1, \dots, Q\} \mid (Bx^{(0)} - b)_q = 0\}$
3 $\quad \tilde{A} := \begin{pmatrix} A \\ B_{\mathcal{Q}} \end{pmatrix}, \quad \tilde{a} := \begin{pmatrix} a \\ b_{\mathcal{Q}} \end{pmatrix}, \quad \tilde{P} := P + |\mathcal{Q}|$
4 $\quad \tilde{F} := I - \tilde{A}^T(\tilde{A}\tilde{A}^T)^{-1}\tilde{A}$
5 $\quad$ for $k := 1, \dots, K$:
6 $\quad\quad \Delta x^{(k-1)} := -\tilde{F}^T \nabla f(x^{(k-1)})$
7 $\quad\quad$ if $||\Delta x^{(k-1)}|| \leq \epsilon$:
8 $\quad\quad\quad$ if $|\mathcal{Q}| = 0$: return $x^{(k-1)}$
9 $\quad\quad\quad \tilde{\lambda} :=$ solve$(\tilde{A}\tilde{\lambda} = -\nabla f(x^{(k-1)}))$
10 $\quad\quad\quad$ if $\tilde{\lambda}_{P+1:\tilde{P}} \geq 0$: return $x^{(k-1)}$
11 $\quad\quad\quad \mathcal{Q} := \mathcal{Q} \setminus \{q\}$ for an arbitrary $q \in \mathcal{Q}$ with $\lambda_q := \tilde{\lambda}_{P+\text{index}(q,\mathcal{Q})} < 0$
12 $\quad\quad\quad$ recompute $\tilde{A}, \tilde{a}, \tilde{P}, \tilde{F}, \Delta x^{(k-1)}$ (= lines 3,4,6)
13 $\quad\quad \mu_{\max}^{(k-1)} := \min\{ \frac{-(Bx^{(k-1)}-b)_q}{(B\Delta x^{(k-1)})_q} \mid q \in \{1, \dots, Q\} \setminus \mathcal{Q}, (B\Delta x^{(k-1)})_q > 0\}$
14 $\quad\quad \mu^{(k-1)} := \mu(f, x^{(k-1)}, \Delta x^{(k-1)}, \mu_{\max}^{(k-1)})$
15 $\quad\quad x^{(k)} := x^{(k-1)} + \mu^{(k-1)}\Delta x^{(k-1)}$
16 $\quad\quad$ if $\mu^{(k-1)} = \mu_{\max}^{(k-1)}$:
17 $\quad\quad\quad \mathcal{Q} := \mathcal{Q} \cup \{q\}$ for an arbitrary $q \in \{1, \dots, Q\} \setminus \mathcal{Q}$ with $\frac{-(Bx^{(k-1)}-b)_q}{(B\Delta x^{(k-1)})_q} = \mu_{\max}^{(k-1)}$
18 $\quad\quad\quad$ recompute $\tilde{A}, \tilde{a}, \tilde{P}, \tilde{F}$ (= lines 3−4)
19 $\quad$ return "not converged"

Gradient Projection / Affine Constraints (ctd.)

where

- $A \in \mathbb{R}^{P \times N}, a \in \mathbb{R}^P$: P affine equality constraints
- $B \in \mathbb{R}^{Q \times N}, b \in \mathbb{R}^Q$: Q affine inequality constraints
- $x^{(0)}$ feasible starting point
- $\mu(\ldots, \mu_{\max})$ step length controller, yielding steplength $\leq \mu_{\max}$
- $\text{index}(q, \mathcal{Q}) := i$ for $q = q_i$ and $\mathcal{Q} = (q_1, q_2, \ldots, q_{\tilde{Q}})$

Remarks

- The projection matrix F does not have to be computed from scratch, every time the active constraint set changes, but can be efficiently updated.

Convergence / Rate of Convergence

- For the gradient projection method, a rate of convergence can be established.
- But the proof is somewhat involved (see [Luenberger and Ye, 2008, ch. 12.5]).

Summary

- Primal methods optimize
  - in the original variables,
  - staying always within the feasible area.
- Backtracking line search can be modified to retain strict inequality constraints.
  - for affine inequality constraints: guaranteed by a maximum stepsize.
- The gradient projection method for affine equality constraints is a modified gradient descent.
  - simply project gradients to the nullspace of the affine constraints.

Summary (2/2)

- Active set methods
  - partition the inequality constraints into active and inactive ones
    - an inequality constraint h_q is active iff $h_q(x) = 0$.
  - add active inequality constraints temporarily to the equality constraints
  - and solve this problem using an optimization method for equality constraints.
  - break away from a random active inequality constraint into whos interior of the feasible area the objective decreases.

- The gradient projection method (for affine equality and inequality constraints) is an active set method that uses the gradient projection method for equality constraints to solve the equality constrained subproblems.

Further Readings

- Primal methods for constrained optimization are not covered by Boyd and Vandenberghe [2004].

- Primal methods often also are called feasible point methods.

- Active set methods:
  - general idea: [Luenberger and Ye, 2008, ch. 12.3]
  - Gradient projection method: [Luenberger and Ye, 2008, ch. 12.4+5], [Griva et al., 2009, ch. 15.4]
  - Reduced gradient method: [Luenberger and Ye, 2008, ch. 12.6+7], [Griva et al., 2009, ch. 15.6]

- Further primal methods not covered here:
  - Frank-Wolfe algorithm / conditional gradient method: [Luenberger and Ye, 2008, ch. 12.1]

References

Stephen Boyd and Lieven Vandenberghe. Convex Optimization. Cambridge University Press, 2004.

Igor Griva, Stephen G. Nash, and Ariela Sofer. Linear and Nonlinear Optimization. Society for Industrial and Applied Mathematics, 2009.

David G. Luenberger and Yinyu Ye. Linear and Nonlinear Programming. Springer, 2008.