# Augmented objective function and methods for constrained optimization

- Penalty method
- Barrier method
- Lagrangian method
- Augmented Lagrangian method

Andersen Ang

Mathématique et recherche opérationnelle, UMONS, Belgium

manshun.ang@umons.ac.be Homepage: angms.science

First draft: August 2, 2017
Last update: December 22, 2020

# Constrained optimization problem

- Unconstrained problem: given a objective function $f : \mathbb{R}^n \to \mathbb{R}$, find $\mathbf{x} \in \mathbb{R}^n$ by solving
$$\min_{\mathbf{x}} f(\mathbf{x}),$$
basically any $\mathbf{x}$ can be a feasible[1] solution.
- Constrained problem
$$\min_{\mathbf{x} \in C} f(\mathbf{x}),$$
not all $\mathbf{x}$ are feasible: $\mathbf{x}$ has to be inside the constraint set $C$.
- If we can characterize $C$ by a set of inequalities and equalities
$$C = \Big\{ \mathbf{x} \mid g_i(\mathbf{x}) = 0,\ h_i(\mathbf{x}) = 0 \Big\},$$
we arrive at the formulation you see in classical textbooks
$$\begin{array}{lll} \min\limits_{\mathbf{x}} & f(\mathbf{x}) & \\ \text{s.t.} & g_i(\mathbf{x}) = 0 & \text{equality constraints} \\ & h_i(\mathbf{x}) \leq 0 & \text{inequality constraints} \end{array}$$

[1] Just may not be optimal.

# Unconstrained and constrained optimization problems

- Unconstrained problem: we can use any optimization method (for example gradient descent) to solve it.
- Constrained problems: gradient descent cannot be applied directly, possible the update moves $\mathbf{x}$ outside the feasible region.
- Modern (new) approach in handling constrained problem: projection gradient or proximal gradient.
- Traditional (old) approach in handling constrained problem: approximate the constrained problem as an unconstrained problem.
  - Lagrangian and Augmented Lagrangian Method
  - Penalty method
  - Barrier method
- Note: being old doesn't means it is always bad, being modern doesn't mean it is always good.

# Reformulation approach

- General idea
  - Reformulate the constrained problem into an unconstrained problem.
  - Solve the unconstrained problem.
  - The sol. of the unconstrained problem is then an approximation of the sol. of the constrained problem.

- A way of reformulation: add a transformed constraint into the objective function

$$\min_{\mathbf{x}} f(\mathbf{x}) \quad \text{s.t.} \quad g(\mathbf{x}) = 0 \quad \rightarrow \quad \min_{\mathbf{x}} \underbrace{f(\mathbf{x}) + \alpha \Phi\big(g(\mathbf{x})\big)}_{F}.$$

  - $f$: objective function
  - $g$: equality constraint
  - $\Phi$: a specially designed function
  - $\alpha \geq 0$: controlling the trade-off between $f$ and $\Phi \circ g$
  - $F$: the augmented objective function

# Why and how of reformulation

- Why reformulate
  The reformulated problem is easier to solve than the original constrained problem.
  - Reformulated problem has no constraint: easier to solve.

- How to reformulate
  Formulate the new problem in such a way that is easier to solve than the original constrained problem.

- What did you paid
  - Introduce more variable.
    The weighting parameter $\alpha$ need tot be chosen properly.
  - Accuracy of solution.
    The solution of the reformulated problem is mostly only an approximation to the solution of the original constrained problem.

# Example: penalty method

- Suppose the problem is

$$\min_{\mathbf{x}} \frac{1}{2}\|\mathbf{x}\|_2^2 \text{ s.t. } \mathbf{A}\mathbf{x} = \mathbf{b}.$$

- A penalty formulation can be

$$\min_{\mathbf{x}} F(\mathbf{x}) := \frac{1}{2}\|\mathbf{x}\|_2^2 + \frac{\alpha}{2}\|\mathbf{A}\mathbf{x} - \mathbf{b}\|_2^2$$

  here $\frac{1}{2}\|\cdot\|$ is the $\Phi$ function.

- This approach is called regularization in machine learning, and $\alpha$ is called the regularization parameter.

- Minimizing $F$ means we find the $\mathbf{x}$ while penalizing the violation of $\mathbf{A}\mathbf{x} - \mathbf{b}$, reflected by the fact that we are also minimizing $\frac{\alpha}{2}\|\mathbf{A}\mathbf{x} - \mathbf{b}\|_2^2$ in $F$.

# Example: penalty method

- Although penalty method is old, it is still useful. In fact many regularized problems in machine learning are instance of penalty method.

- Suppose the problem is

$$(\mathcal{P}_0) \;\; : \;\; \min_{\mathbf{x}} \|\mathbf{A}\mathbf{x} - \mathbf{b}\|_2^2 \text{ s.t. } \|\mathbf{x}\|_1 \leq \epsilon.$$

  Such problem is hard to solve since it is hard to project onto the constraint $\|\mathbf{x}\|_1 \leq \epsilon$.

- A regularized problem

$$(\mathcal{P}_1) \;\; : \;\; \min_{\mathbf{x}} F(\mathbf{x}) := \|\mathbf{A}\mathbf{x} - \mathbf{b}\|_2^2 + \lambda \|\mathbf{x}\|_1.$$

- For each $\epsilon$ in $(\mathcal{P}_0)$, there will be a $\lambda$ in $(\mathcal{P}_1)$ that their sol. are equal.

# Example: barrier method

- Suppose the problem is to minimize a scalar quadratic function over nonnegative constraint

$$\min_x ax^2 + bx + c \text{ s.t. } x \geq 0.$$

- A log-barrier formulation can be

$$\min_x F(\mathbf{x}) := ax^2 + bx + c - \alpha \log x$$

  here $-\log x$ is the log-barrier: as $x \to 0^+$, the log term goes to $\infty$.

- Note: barrier method is practically not good for nonnegativity constraint.

# Example: Lagrangian method

- Suppose the problem

$$\begin{array}{lll} \min\limits_{\mathbf{x}} & f(\mathbf{x}) & \\ \text{s.t.} & g_i(\mathbf{x}) = 0 & \text{equality constraints} \\ & h_i(\mathbf{x}) \leq 0 & \text{inequality constraints} \end{array}$$

- Lagrangian method / KKT method

$$L(\mathbf{x}, \lambda_1, \lambda_2, \ldots, \nu_1, \nu_2, \ldots) = f(\mathbf{x}) + \sum_i \lambda_i g_i(\mathbf{x}) + \sum_j \nu_i h_i(\mathbf{x}),$$

  $\lambda_i$: Lagrangian multipliers, and $\nu_i$: KKT multipliers.

- Compact notation: $\boldsymbol{\lambda} = [\lambda_1, \lambda_2, \ldots]^\top$, $\boldsymbol{\nu} = [\nu_1, \nu_2, \ldots]^\top$, $\mathbf{g}(\mathbf{x}) = [g_1(\mathbf{x}), g_2(\mathbf{x}), ...]$ , $\mathbf{h}(\mathbf{x}) = [h_1(\mathbf{x}), h_2(\mathbf{x}), ...]$ ,

$$L(\mathbf{x}, \boldsymbol{\lambda}, \boldsymbol{\nu}) = f(\mathbf{x}) + \langle \boldsymbol{\lambda}, \mathbf{g}(\mathbf{x}) \rangle + \langle \boldsymbol{\nu}, \mathbf{h}(\mathbf{x}) \rangle.$$

- The Lagrangian method consists of solving $L(\mathbf{x}, \boldsymbol{\lambda}, \boldsymbol{\nu})$ with respect to the KKT conditions.

# Example: augmented Lagrangian method (ALM)

- Suppose the problem (in compact notation for constraints)

$$(\mathcal{P}_0) \ : \ \min_{\mathbf{x}} f(\mathbf{x}) \text{ s.t. } \mathbf{g}(\mathbf{x}) = 0, \mathbf{h}(\mathbf{x}) \leq 0.$$

- ALM is actually a penalty method: it consider the problem

$$(\mathcal{P}_1) \ : \ \min_{\mathbf{x}} f(\mathbf{x}) + \frac{\alpha}{2}\|\mathbf{g}(\mathbf{x})\|_2^2 + \frac{\beta}{2}\|\mathbf{h}(\mathbf{x})\|_2^2 \text{ s.t. } \mathbf{g}(\mathbf{x}) = 0, \mathbf{h}(\mathbf{x}) \leq 0.$$

- The Lagrangian of $(\mathcal{P}_1)$ is

$$L_{\alpha,\beta}(\mathbf{x}, \boldsymbol{\lambda}, \boldsymbol{\nu}) = f(\mathbf{x}) + \frac{\alpha}{2}\|\mathbf{g}(\mathbf{x})\|_2^2 + \frac{\beta}{2}\|\mathbf{h}(\mathbf{x})\|_2^2 + \langle \boldsymbol{\lambda}, \mathbf{g}(\mathbf{x}) \rangle + \langle \boldsymbol{\nu}, \mathbf{h}(\mathbf{x}) \rangle.$$

  it is called the augmented Lagrangian.

- What's new in Augmented Lagrangian: the quadratic terms. Their function is to improve the convexity condition of $L$.

- ALM solve $(\mathcal{P}_1)$ by minimizing $L_{\alpha,\beta}(\mathbf{x}, \boldsymbol{\lambda}, \boldsymbol{\nu})$ with respect to $\mathbf{x}$, then update $\boldsymbol{\lambda}$ and $\boldsymbol{\nu}$.

# Last page - summary

Conceptual introduction to augmented objective function

- Penalty method
- Barrier method
- Lagrangian method
- Augmented Lagrangian method

Details of these methods: see other documents.

End of document