Fast review of augmented cost function

Andersen Ang

ECS, Uni. Southampton, UK

andersen.ang@soton.ac.uk

Homepage angms.science

Version: April 1, 2023
First draft: August 2, 2017

Content

Constrained optimization problem

- Given $f : \mathbb{R}^n \to \mathbb{R}$.

- Unconstrained optimization $\underset{\boldsymbol{x}}{\operatorname{argmin}}\ f(\boldsymbol{x})$ all $\boldsymbol{x}$ feasible

- Constrained problem $\underset{\boldsymbol{x}\in\mathcal{C}}{\operatorname{argmin}}\ f(\boldsymbol{x})$ not all $\boldsymbol{x}$ feasible

- If we can characterize $\mathcal{C}$ by a set of inequalities and equalities

$$\mathcal{C} = \left\{\boldsymbol{x} \,\middle|\, g_i(\boldsymbol{x}) = 0,\ h_i(\boldsymbol{x}) = 0\right\},$$

  we arrive at the a textbook formulation

$$\begin{aligned} \underset{\boldsymbol{x}}{\operatorname{argmin}} \quad f(\boldsymbol{x}) \quad \text{s.t.} \quad & g_i(\boldsymbol{x}) = 0 \quad \text{equality constraints} \\ & h_i(\boldsymbol{x}) \leq 0 \quad \text{inequality constraints} \end{aligned}$$

Solution approach for constrained problem

- Modern approach: projection / proximal gradient.
- Traditional approach: reformulation
  - Lagrangian and Augmented Lagrangian Method
  - Penalty method
  - Barrier method
- Being old doesn't means it is always bad, being modern doesn't mean it is always good.

Reformulation approach

- 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.

- Reformulation: add a transformed constraint to the cost

$$\underset{\boldsymbol{x}}{\operatorname{argmin}}\ f(\boldsymbol{x}) \ \text{ s.t. } \ g(\boldsymbol{x}) = 0 \quad \longrightarrow \quad \underset{\boldsymbol{x}}{\operatorname{argmin}}\ \underbrace{f(\boldsymbol{x}) + \alpha\Phi\big(g(\boldsymbol{x})\big)}_{F}.$$

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

Why and how of reformulation

- Why reformulate
  The reformulated problem is *easier* to solve than the original one
- How to reformulate
  Formulate a new problem that is *easier* to solve than the original one
- What did you paid
  - Introduce more variable.
    The weighting parameter α need to be chosen properly.
  - Accuracy of solution.
    The sol. of the reformulated problem may only be an approximation sol. of the original one.

Example: penalty method

- Original problem
$$\operatorname*{argmin}_{\boldsymbol{x}} \frac{1}{2}\|\boldsymbol{x}\|_2^2 \text{ s.t. } \boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}.$$
- A penalty formulation
$$\operatorname*{argmin}_{\boldsymbol{x}} F(\boldsymbol{x}) := \frac{1}{2}\|\boldsymbol{x}\|_2^2 + \frac{\alpha}{2}\|\boldsymbol{A}\boldsymbol{x} - \boldsymbol{b}\|_2^2$$
  here $\frac{1}{2}\|\cdot\|$ is the Φ in the previous slide.
- This approach is called *regularization* in machine learning, α is the regularization parameter.
- Minimizing F means we find $\boldsymbol{x}$ while penalizing the violation of $\boldsymbol{A}\boldsymbol{x} \neq \boldsymbol{b}$, because $\frac{\alpha}{2}\|\boldsymbol{A}\boldsymbol{x} - \boldsymbol{b}\|_2^2 > 0$ if $\boldsymbol{A}\boldsymbol{x} \neq \boldsymbol{b}$ will increase F, and we want F small.
- <u>Penalty method is (very) old[1] but st</u>ill useful.

[1]As old as in the 1940s

Many regularized problems in machine learning are instances of penalty method

- A machine learning problem

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

- A regularized form

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

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

- Another approach (less intuitive)

$$(\mathcal{P}_2) \;:\; \min_{\boldsymbol{x}} G(\boldsymbol{x}) := \|\boldsymbol{A}\boldsymbol{x} - \boldsymbol{b}\|_2^2 + i_{\|\cdot\|_1 \leq \epsilon}(\boldsymbol{x}).$$

  Solving $\mathcal{P}_2$ involves the proximal operator, duality, conjugate and Moreau decomposition, which is less intuitive.

Example: barrier method

- Consider minimizing a scalar quadratic function over nonnegative constraint

$$\operatorname*{argmin}_{x} ax^2 + bx + c \,\text{s.t.}\, x \geq 0.$$

- A log-barrier formulation

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

here $-\log x$ is the log-barrier: as $\lim\limits_{x \to 0^+} -\log x = \infty$.

Example: Lagrangian method

- Suppose the problem is

$$\operatorname*{argmin}_{\boldsymbol{x}} f(\boldsymbol{x}) \quad \text{s.t.} \quad \begin{array}{ll} g_i(\boldsymbol{x}) = 0 & \text{equality constraints} \\ h_i(\boldsymbol{x}) \leq 0 & \text{inequality constraints} \end{array}$$

- Lagrangian method / KKT method

$$\begin{aligned} \mathcal{L}(\boldsymbol{x}, \lambda_1, \lambda_2, \ldots, \nu_1, \nu_2, \ldots) &= f(\boldsymbol{x}) + \sum_i \lambda_i g_i(\boldsymbol{x}) + \sum_j \nu_i h_i(\boldsymbol{x}) \\ &= f(\boldsymbol{x}) + \langle \boldsymbol{\lambda}, \boldsymbol{g}(\boldsymbol{x}) \rangle + \langle \boldsymbol{\nu}, \boldsymbol{h}(\boldsymbol{x}) \rangle \end{aligned}$$

  - λ_i: Lagrangian multipliers
  - ν_i: KKT multipliers
  - $\boldsymbol{\lambda} = [\lambda_1, \ldots]^\top$, $\boldsymbol{\nu} = [\nu_1, \ldots]^\top$, $\boldsymbol{g}(\boldsymbol{x}) = [g_1(\boldsymbol{x}), \ldots]$, $\boldsymbol{h}(\boldsymbol{x}) = [h_1(\boldsymbol{x}), \ldots]$

- The Lagrangian method = solving a saddle point problem associated with $\mathcal{L}$

$$\operatorname*{argmin}_{\boldsymbol{x}} \operatorname*{argmax}_{\boldsymbol{\lambda}, \boldsymbol{\nu} \geq \boldsymbol{0}} \mathcal{L}(\boldsymbol{x}, \boldsymbol{\lambda}, \boldsymbol{\nu})$$

Example: augmented Lagrangian method (ALM)

- Suppose the problem

$$(\mathcal{P}_0) \ : \ \underset{\boldsymbol{x}}{\operatorname{argmin}} \ f(\boldsymbol{x}) \ \text{ s.t. } \ \boldsymbol{g}(\boldsymbol{x}) = \mathbf{0}.$$

- ALM is a penalty method: it consider the problem

$$(\mathcal{P}_1) \ : \ \underset{\boldsymbol{x}}{\operatorname{argmin}} \ f(\boldsymbol{x}) + \frac{\alpha}{2}\|\boldsymbol{g}(\boldsymbol{x})\|_2^2 \ \text{ s.t. } \ \boldsymbol{g}(\boldsymbol{x}) = \mathbf{0}.$$

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

$$\mathcal{L}_\alpha(\boldsymbol{x}, \boldsymbol{\lambda}) = f(\boldsymbol{x}) + \frac{\alpha}{2}\|\boldsymbol{g}(\boldsymbol{x})\|_2^2 + \langle \boldsymbol{\lambda}, \boldsymbol{g}(\boldsymbol{x}) \rangle.$$

- $\mathcal{L}_\alpha(\boldsymbol{x}, \boldsymbol{\lambda})$ it is called the augmented Lagrangian. The quadratic term improves the convexity condition of L.

Last page - summary

Conceptual introduction to augmented cost function

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

Details of these methods: see other documents.

End of document