

Constrained Optimization

Monday, September 30, 2024

9:37 AM

Ref: [convex-optimization-class/TypedNotes/lecture_notes.pdf at master · stephenbeckr/convex-optimization-class \(github.com\)](https://github.com/stephenbeckr/convex-optimization-class) and Nocedal and Wright textbook 2005

$$\min_{\mathbf{w}} f(\mathbf{w}) \quad \text{s.t.} \quad h_j(\mathbf{w}) = 0, \quad j=1, \dots, m$$

inequality constraints
even more complicated
in some ways

no longer take $\nabla f(\mathbf{w}) = 0$... more complicated.

One class of methods is to incorporate h_j constraints into our model m_k

- Common approach when h is simple
(eg. enforce $\mathbf{w} \geq 0$)
- Less common if h is complicated or implicit
(eg. \mathbf{w} solves a PDE)

Another class is to reformulate

Simplest: the penalty method

Very common (eg. PINNs)

due to simplicity,

not due to effectiveness.

Solve

$$\min_{\mathbf{w}} f(\mathbf{w}) + \frac{1}{2} \sum_{j=1}^m \mu_j \cdot h_j(\mathbf{w})^2$$

for $\mu_j > 0$
often $\mu_j \equiv \mu$

Idea: set μ very large

(Slight improvement: solve for moderate μ , then increase μ and re-solve, using your old value of \mathbf{w} as a "warm-start")

Disadvantages:

- ① it need not work
- ② it can get harder to solve as $\mu \rightarrow \infty$
- ③ in practice, sometimes solve for a single μ , so it's not really even solving the problem.

Penalty method

Monday, September 30, 2024 6:14 PM

Ex: ill-conditioning of penalty method as $\mu \rightarrow \infty$

$$\min_w \frac{1}{2} w^T H w \text{ s.t. } Aw = b \quad \text{s.t. } a_i^T w = b_i \text{ so } h_i(w) = a_i^T w - b_i;$$

$$\text{as penalty method, } \min_w \frac{1}{2} w^T H w + \mu/2 \|Aw - b\|_2^2$$

so solution is

$$w = \underbrace{(H + \mu A^T A)^{-1} A^T b}_{\text{condition number } \geq \mu}$$

\Rightarrow iterative solvers will converge slowly

Ex: penalty method doesn't always work

$$\min_{w \in \mathbb{R}^n} w^3 \text{ s.t. } w=1$$

Solution is $w=1$

$$\left. \begin{array}{l} \min_w w^3 + \mu/2 (w-1)^2 \\ w^3 \\ w^2 \end{array} \right\}$$

Solution is $w = -\infty$ for all μ

Thm (Nocedal & Wright)

i.e. true minimizer, not just stationary pt.

For equality constraints, if w_k solves the penalized problem

with μ_k , then if (w_k) has a limit point w^* ,

this limit point solves the original (constrained) problem

Inequality constraints

e.g. $\min f(w) \text{ s.t. } h(w) \leq 0$

penalized version is $\min f(w) + \mu/2 \|h(w)\|_+$

where $\|\alpha\|_+ := \max(0, \alpha)$ is the positive part
(aka ReLU!)

Lagrange Multipliers

Monday, September 30, 2024 6:14 PM

Connect w/ what you see in calculus ref. Stewart "Essential Calc." § 11.8

$$\min_{w \in \mathbb{R}^2} f(w) \text{ s.t. } g(w) = c, \quad f, g \in C^1$$

Thm If w^* is a minimizer of the above problem and $\nabla g(w^*) \neq 0$
then $\exists \lambda \in \mathbb{R}$ s.t. $\nabla f(w^*) = \lambda \nabla g(w^*)$



ref. Boyd & Vandenberghe "convex optimization" ~2004

General Case: dual variables and the KKT conditions

$$\min_{w \in \mathbb{R}^n} f(w) \text{ st. } \begin{aligned} g_j(w) &= 0 & j = 1, \dots, m_E & \text{Equality constraints} \\ h_j(w) &\leq 0 & j = 1, \dots, m_I & \text{Inequality constraints} \end{aligned}$$

Define the Lagrangian

$$\lambda \in \mathbb{R}^{m_I}, \nu \in \mathbb{R}^{m_E}$$

$$L(w, \lambda, \nu) := f(w) + \sum_{j=1}^{m_I} \lambda_j h_j(w) + \sum_{j=1}^{m_E} \nu_j g_j(w)$$

If $\lambda \geq 0$ this is a lower bound on optimal value
of original ("primal") problem

("dual" problem is then
finding largest
lower bound)

KKT conditions

$$\textcircled{1} \text{ "Stationarity" } \nabla_w L = 0$$

← i.e. $w = \arg \min_w L(w, \lambda, \nu)$

$$\textcircled{2} \text{ Feasibility: } g_j(w) = 0, h_j(w) \leq 0$$

ignore if
no inequality
constraints

$$\textcircled{3} \text{ Dual feasibility } \lambda_j \geq 0 \quad \forall j$$

$$\textcircled{4} \text{ Complementary Slackness } \lambda_j h_j(w) = 0 \quad \forall j$$

"Thm" Assuming derivatives exist, etc., the KKT conditions

are necessary. For Convex problems (f, h_j convex, g_j affine)
they are also sufficient under mild constraint qualifications.

i.e. w^* is a global minimizer (of the constrained problem)

⇒ w^* satisfies the KKT conditions

↑
"↔" Sometimes

Augmented Lagrangian

Monday, September 30, 2024

6:14 PM

Refs: Bertsekas' "Convex Optimization Algorithms" §5.2
or Boyd et al.'s ADMM monograph from ~2011

$$\min f(\omega) \quad \text{s.t. } h_j(\omega) = 0 \quad j=1, \dots, m$$

equivalent to this "Augmented" problem

$$\min f(\omega) + \mu/2 \sum_{j=1}^m h_j(\omega)^2 \quad \text{s.t. } h_j(\omega) = 0$$

Now form the Lagrangian for this augmented problem

$$\mathcal{L}_\mu(\omega, \gamma) = f(\omega) + \mu/2 \sum_{j=1}^m h_j(\omega)^2 + \sum_{j=1}^m \gamma_j h_j(\omega)$$

Aug. Lagr. algo :

- initialize $\gamma^{(0)}$
- for $k = 0, 1, 2, \dots$
- $\omega^{(k+1)} = \underset{\omega}{\operatorname{argmin}} \mathcal{L}_\mu(\omega, \gamma^{(k)})$ // primal update
- $\gamma_j^{(k+1)} = \gamma_j^{(k)} + \mu h_j(\omega^{(k+1)})$ for $j=1, \dots, m$ // dual update
- optional: increase μ

if solving iteratively, warm-start using $\omega^{(k)}$

aka "Method of Multipliers"

Bertsekas 1970's

(\neq ADMM, "Alternating Direction Method of Multipliers")
though it is related

Idea: if we knew true γ^* from KKT equations, we could

just solve unconstrained problem $\min_{\omega} \mathcal{L}_\mu(\omega, \gamma^*)$ (for any μ)

$\mu = 0$ is "dual ascent"

$\mu > 0$ regularizes, makes dual problem nice.

Theory isn't super strong but good practical performance

Unlike the penalty method, you don't need $\mu \rightarrow \infty$

... so fewer ill-conditioning issues

with inequality constraints...

gets tricky, see Nocedal & Wright, Software like LANCELOT, MINOS...

Alternatives (not in detail)

Monday, September 30, 2024 6:14 PM

Other ways to deal with constraints:

- Put them into the model m_x

Simple constraints

$$\text{eg. } w \geq 0$$

$$1 \leq w \leq u$$

$$\|w\| \leq \varepsilon$$

} projected gradient descent family

complicated constraints

usually (iteratively) simplify them

i.e. linearize via Taylor Series

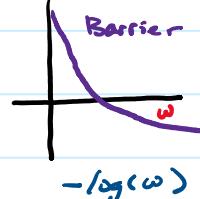
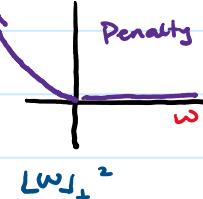
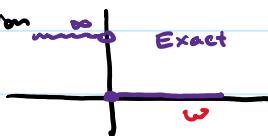
sequential quadratic programming

sequential convex programming family

- Penalty methods used idea of a penalty function

Instead, use a barrier function

$$w \geq 0$$



Interior point method family

- Misc. methods

Some overlap w/ nonsmooth techniques

- Tricks, reparameterization

ex: $w \in \mathbb{R}^d$, want $w \geq 0$, reparameterize $w = v^2$

ex: $W \in \mathbb{R}^{d \times d}$, want $W \geq 0$ (i.e. symmetric positive semidefinite)
reparameterize $W = V \cdot V^T$

ex: $Aw = b$, reparameterize $w = w_p + N \cdot v$ \nwarrow variable
 \nearrow Fixed

where w_p is any (fixed)

solution to $Aw_p = b$ "particular solution" in ODE terminology

and columns of N form a basis for the null space of A

i.e. $\text{col}(N) = \text{ker}(A)$

Can find via SVD or
Standard linear algebra

$$\text{so } A \cdot (w_p + N \cdot v) = Aw_p + ANv = b + 0 = b \checkmark$$