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) and Nocedal and Wright textbook 2005

$\min_{w} f(w)$ s.t. $h_j(w) = 0$, $j = 1, \ldots, m$ *

* inequality constraints even more complicated in some ways

no longer take $\nabla f(w) = 0$ ... 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 $w \geq 0$)
- Less common if h is complicated or implicit (eg. 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

$$\min f(w) + \frac{1}{2} \sum_{j=1}^{m} \mu_j \cdot h_j(w)^2 \quad \text{for } \mu_j > 0$$

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 w as a "warm-start")

Disadvantages:
1. it need not work
2. it can get harder to solve as $\mu \to \infty$
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 $\mu \to \infty$

$$\min_{\omega} \ \tfrac{1}{2}\omega^T H \omega \quad \text{s.t.} \quad A\omega = b$$ ← i.e. $a_i^T\omega = b_i$ so $h_i(\omega) = a_i^T\omega - b_i$

as penalty method, $\min_{\omega} \ \tfrac{1}{2}\omega^T H\omega + \mu/2 \, \|A\omega - b\|_2^2$

So solution is

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

$\Rightarrow$ iterative solvers will converge slowly

Ex: penalty method doesn't always work

$$\min_{\omega \in \mathbb{R}^1} \ \omega^3 \quad \text{s.t. } \omega = 1 \qquad \Big\} \qquad \min_{\omega} \ \omega^3 + \mu/2 \, (\omega - 1)^2$$

Solution is $\omega = 1$

ω^3 ω^2

Solution is $\omega = -\infty$ for all μ

Thm (Nocedal + Wright)

For equality constraints, if ω_k solves the penalized problem with μ_k, and $\lim_{k\to\infty} \mu_k = \infty$, then if (ω_k) has a limit point ω^*, this limit point solves the original (constrained) problem

(solves: i.e. true minimizer, not just stationary pt.)

Inequality constraints

e.g. $\min f(\omega)$ s.t. $h(\omega) \leq 0$

penalized version is $\min f(\omega) + \mu/2 \, \lfloor h(\omega) \rfloor_+^2$

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

Lagrange Multipliers

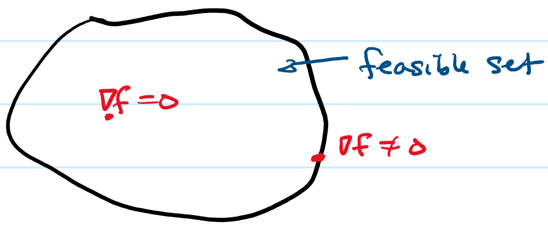
Monday, September 30, 2024 6:14 PM

Connect w/ what you see in calculus

ref. Stewart "Essential Calc." §11.8

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

Thm If w^* is a minimizer of the above problem and $\nabla g(w^*) \neq 0$ then $\exists\ \lambda \in \mathbb{R}$ s.t. $\underbrace{\nabla f(w^*) = \lambda \nabla g(w^*)}_{\text{necessary (but not sufficient)}}$



ref. Boyd + Vandenberghe "convex optimization" ~2004

General case: dual variables and the KKT conditions

$$\min_{w \in \mathbb{R}^n} f(w) \quad \text{s.t.} \quad \begin{array}{ll} g_j(w) = 0 & j = 1, \dots, m_E \quad \text{Equality constraints} \\ h_j(w) \le 0 & j = 1, \dots, m_I \quad \text{Inequality constraints} \end{array}$$

Define the **Lagrangian** $\lambda \in \mathbb{R}^{m_I}$, $\nu \in \mathbb{R}^{m_E}$

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

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

1. "Stationarity" $\nabla_w \mathcal{L} = 0$ ← i.e. $w = \operatorname{argmin}_w \mathcal{L}(w, \lambda, \nu)$
2. feasibility: $g_j(w) = 0$, $h_j(w) \le 0$
3. dual feasibility $\lambda_j \geq 0 \ \forall j$
4. complementary slackness $\lambda_j h_j(w) = 0 \ \forall j$

ignore if no inequality constraints { (3), (4) }

"**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)

$\Rightarrow$ w^* satisfies the KKT conditions

"$\Leftarrow$" 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

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

equivalent to this "Augmented" problem

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

Now form the Lagrangian for this augmented problem

$$\mathcal{L}_\mu(w, \nu) = f(w) + \mu/2 \sum_{j=1}^{m} h_j(w)^2 + \sum_{j=1}^{m} \nu_j h_j(w)$$

Aug. Lagr. algo: initialize $\nu^{(0)}$

for $k = 0, 1, 2, \dots$

$$w^{(k+1)} = \operatorname{argmin}_w \mathcal{L}_\mu(w, \nu^{(k)}) \quad \text{// primal update}$$

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

$$\nu_j^{(k+1)} = \nu_j^{(k)} + \mu h_j(w^{(k+1)}) \quad \text{for } j = 1, \dots, m$$

// dual update

optional: increase μ

aka "Method of Multipliers"
Bertsekas 1970's

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

Idea: if we knew true ν^* from KKT equations, we could just solve unconstrained problem $\min_w \mathcal{L}_\mu(w, \nu^*)$ (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 $\mu \to \infty$

... 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 <u>model</u> m_k

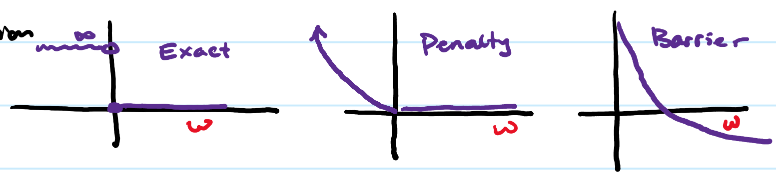
  Simple constraints
  e.g. $w \geq 0$
  $l \leq w \leq u$
  $\|w\| \leq \varepsilon$
  } <u>projected</u> gradient descent family

  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 <u>penalty</u> function

  Instead, use a <u>barrier</u> function

  $w \geq 0$



$\lfloor w \rfloor_+^2$ $\qquad$ $-\log(w)$

  Interior point method family

- Misc. methods

  Some overlap w/ nonsmooth techniques

- Tricks, reparameterization

  ex: $w \in \mathbb{R}^1$, want $w \geq 0$, reparameterize $w = v^2$

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

  ex: $Aw = b$, reparameterize $w = w_p + N \cdot v$ ← variable
  ↖ ↗ Fixed
  where w_p is <u>any</u> (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. $\mathrm{col}(N) = \ker(A)$ Can find via SVD or standard linear algebra

  so $A \cdot (w_p + N \cdot v) = \underbrace{Aw_p}_{b} + \underbrace{ANv}_{0} = b$ ✓