2/2 recitation

- questions
- quadratic forms
- KKT conditions
- QPAL solver

Merit fx

There were some questions about merit functions. In an iterative solver, we are computing a sequence of xk's that lead us to a solution. In order to ensure we are headed in the right direction, we want to make sure each xk is better than the previous. We use whatever method we want to (gradient descent, Newton, etc) to compute a delta_x, and a merit function is what we use to determine how far we should step in that direction. We want the merit (phi) to decrease with the step, so we use a linesearch to find an alpha that satisfies phi(xk + alpha*delta_x) < phi(xk).

$$x \,,\ \Delta x$$

$$x_{k+1} = x_k + \alpha \Delta x$$

$$\phi(x_k + \alpha \Delta x) < \phi(x_k)$$

Here is an example of a merit function, where we just have the cost + a variation of constraint violation.

if $\min_x f(x)$ s.t. $c(x) = 0$ $\qquad \phi(x) = f(x) + \rho \|c(x)\|_1$

---

quadratic forms

Many of the objective functions we use in control can be reduced to a quadratic form. This is nice for us because it is simple to take derivatives of. Also, if it's an objective function, we don't care about constant terms, they don't change the minimizing argument.

$$f(x) = \frac{1}{2} x^T Q x + q^T x + \cancel{r}$$

we don't care about constants when minimizing

3 matrix manipulations

1. $(ABC)^T = C^T B^T A^T$
2. $x^T A b = (x^T A b)^T = b^T A^T x$
3. $(Ax - b)^T = x^T A^T - b^T$

* Matrix Cookbook

Here are 3 things to keep in mind when manipulating matrix equations.

LS

$$\frac{1}{2}\|Ax - b\|_2^2 = \frac{1}{2}(Ax-b)^T(Ax-b)$$

$$= \frac{1}{2}(x^T A^T - b^T)(Ax - b)$$

$$= \frac{1}{2}\left(x^T A^T A x - b^T A x - \underbrace{x^T A^T b}_{②} + b^T b\right)$$

constant, ignore

$$= \frac{1}{2}\left(x^T A^T A x - b^T A x - b^T A x + b^T b\right)$$

$$= \frac{1}{2}\left(x^T A^T A x - 2\underbrace{b^T A}_{①} x + b^T b\right)$$

$$= \frac{1}{2} x^T \underbrace{A^T A}_{Q} x - \underbrace{(A^T b)}_{q}{}^T x + \underbrace{\frac{1}{2} b^T b}_{r}$$ Constant

Here we put the classic least squares objective into a quadratic form.

taking derivs in a quadratic form

$$f(x) = \frac{1}{2} x^T Q x + q^T x + r$$

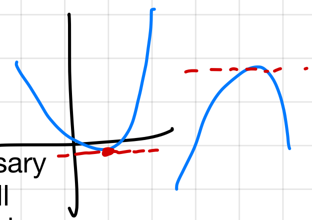
$$\nabla_x f = Qx + q$$

$$\nabla_x^2 f = Q$$

Gradient and hessian of a quadratic.

KKT Conditions $\quad$ 1^{st}-order neccessary conditions $\quad$ if convex, its sufficient

① Unconstrained: $\min_x \; f(x)$



$$\nabla_x f(x) = 0$$

KKT conditions are necessary optimality conditions for all optimization problems, and sufficient for convex problems. If we find a solution that satisfies the KKT conditions for a convex problem, we have found the global solution.

Least squares example

$$\nabla_x \left(\frac{1}{2} x^T A^T A x - (A^T b)^T x + \frac{1}{2} b^T b \right) = A^T A x - A^T b = 0$$

$$\boxed{A^T A x = A^T b}$$

② eq - constrained

$$\boxed{\begin{aligned} \min_x \quad & f(x) \\ \text{s.t.} \quad & c(x) = 0 \end{aligned}}$$

primal var: $x \in \mathbb{R}^n$

dual var: $\lambda \in \mathbb{R}^m$, $\quad c(x) \in \mathbb{R}^m$

Dual variable is the size of the equality constraint.

Lagrangian $\quad L(x, \lambda) = f(x) + \lambda^T c(x)$

Stationarity: $\quad \nabla_x L(x, \lambda) = \nabla_x f(x) + \left(\frac{\partial c}{\partial x}\right)^T \lambda = 0$

primal feasibility: $\quad c(x) = 0$

These two equations are the KKT conditions for an equality constrained optimization problem. Notice how they are equations that we can use Newton on to find a primal-dual solution.

③ ineq constrained problem

Here are the total KKT conditions for all problems, the first two examples were special cases. The convention used here (wrt the sign of the inequality and the dual feasibility constraint) is consistent and will always work for you.

$$\min_x \quad f(x)$$
$$\text{s.t.} \quad c(x) = 0 \qquad \lambda$$
$$g(x) \leq 0 \qquad \mu$$

dual

Lagrangian

$$L(x, \lambda, \mu) = f(x) + \lambda^T c(x) + \mu^T g(x)$$

KKT

Stationarity: $\nabla_x L = \nabla_x f(x) + \left(\frac{\partial c}{\partial x}\right)^T \lambda + \left(\frac{\partial g}{\partial x}\right)^T \mu = 0$

primal feasibility: $c(x) = 0$

$g(x) \leq 0$

dual feasibility $\mu \geq 0$

Complimentarity slackness

① $\mu_i \cdot g(x)_i = 0 \quad \forall i$ "for all"

② $\mu \odot g(x) = 0$ element wise product

1 and 2 are the same but just have different notation. 3 is true, but I feel like it's less useful than 1 and 2 for understanding complimentarity.

③ $\mu^T g(x) = 0$ not super useful

$$g(x) = \begin{bmatrix} -3 \\ 0 \\ -2 \\ 0 \end{bmatrix} \qquad \mu = \begin{bmatrix} 0 \\ \sim \\ 0 \\ \sim \end{bmatrix}$$

inactive

active

Here is an example of complimentarity.

$$g(x) \odot \mu = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}$$

QPAL

$\text{eig}(Q) \geq 0,\ Q \in S_+$ means its convex

$$\min_x \quad \frac{1}{2} x^T Q x + g^T x$$

$$\text{s.t.} \quad Ax - b = 0$$

$$Gx - h \leq 0$$

$$L(x, \lambda, \mu) = \frac{1}{2} x^T Q x + g^T x + \lambda^T (Ax - b) + \mu^T (Gx - h)$$

KKT

Stationarity $\quad \nabla_x L = Qx + g + A^T \lambda + G^T \mu = 0$

primal $\quad Ax - b = 0$

feasibility $\quad Gx - h \leq 0$

dual feasibility $\quad \mu \geq 0$

compl. slackness $\quad (Gx - h) \odot \mu = 0$

Here are the KKT conditions for a convex QP. Again, any primal-dual solution that satisfies these is a the global minima.

Augmented Lagrangian

$$L_\rho(x, \lambda, \mu, \rho) = L_\rho(x, \lambda, \mu) + \underbrace{\frac{\rho}{2} \|Ax - b\|_2^2}_{\text{eq}} + \underbrace{\frac{1}{2} (Gx - h)^T I_\rho (Gx - h)}_{\text{ineq},\ I_\rho \text{ is mask}}$$

$I_\rho \in S^m$, μ, $Gx-h$ both $\in \mathbb{R}^m$

```
for i := 1:m
    if (Gx-h)_i < 0 and μ_i = 0
        I_ρ[i,i] = 0
    else
        I_ρ[i,i] = ρ
```

This is how to make the mask matrix.

QPAC

init $x = 0, \lambda = 0, \mu = 0, \rho = 1, \phi = 10$

for $i = 1$: max_AL_iters

① solve $\min_x L_\rho(x, \lambda, \mu, \rho)$

with Newton's method. It may take multiple Newton iterations. don't stop until $\nabla_x L_\rho = 0$

$$x_{k+1} = x_k - (\nabla_x^2 L_\rho)^{-1} \nabla_x L_\rho \quad \text{until } \nabla_x L_\rho = 0$$

② update duals

$$\lambda = \lambda + \rho(Ax - b)$$

$$\mu = \max(0, \mu + \rho(Gx - h))$$

③ update ρ

$$\rho = \rho \cdot \phi$$

The biggest mistake people made last year was not doing (1) correctly. You need to solve the unconstrained minimization problem of the Augmented Lagrangian to convergence. A lot of people were just taking one Newton step and moving on to step (2). Do not move to step (2) until you have solved (1).

④ Check Convergence

$$\text{eq\_vio} = \|Ax - b\|_\infty$$

$$\text{ineq\_vio} = \max(0, \text{maximum}(Gx - h))$$

Julia