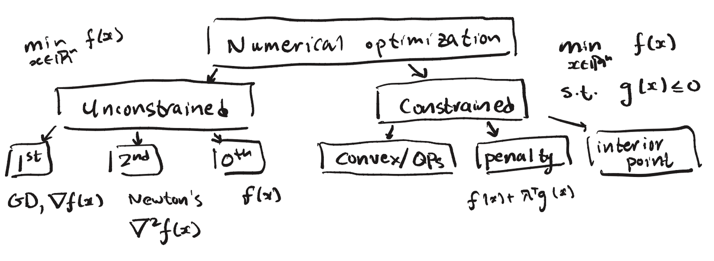
2/24: Penalty methods & AL

Announcements

- HW 2 out, due 3/5 (note: leaderboard in FLOPs)
- project feedback @ end of week
- slight reschedule

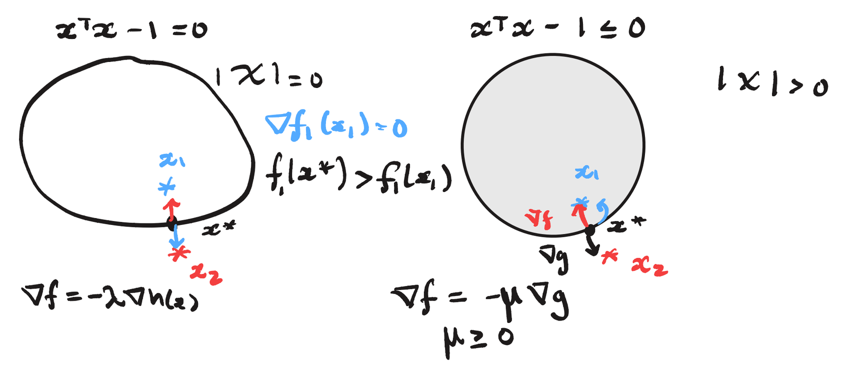
Today:

- KKT discussion
- penalty methods
- Augmented Lagrangian



Recall:

$$\min_x \; f(x)$$
$$\text{s.t. } h_i(x) = 0, \; i = 1, \dots, \ell$$
$$g_j(x) \le 0, \; j = 1, \dots, m$$



Def1 (KKT Conditions)

A point x^* satisfies KKT if $\exists \, \lambda^*, \mu^*$, satisfying:

1. Stationarity: $\nabla f(x^*) + \sum_{i=1}^{\ell} \lambda_i \nabla h_i(x^*) + \sum_{j=1}^{m} \mu_j \nabla g_j(x^*) = 0$ $\left(\nabla_x \mathcal{L}(x^*, \lambda^*, \mu^*) = 0 \right)$
2. Feasibility: $h_i(x^*) = 0 \quad \forall i$, $\left(\nabla_\lambda \mathcal{L}(x^*, \lambda^*, \mu^*) = 0 \right)$
   $g_j(x^*) \le 0 \quad \forall j$
3. Dual feasibility: $\mu_j^* \ge 0$
4. Complementarity $\mu^* \odot g(x^*) = 0$
   $\mu_j^* = 0$ or $g_j(x^*) = 0$

Fact: For smooth constraints, KKT are necessary optimality conditions

---

Recall: Lagrangian "pricing in"

$$\mathcal{L}(x, \lambda, \mu) = f(x) + \lambda^T h(x) + \mu^T g(x)$$

Fact: The original constrained problem is equivalent to:

$$\min_{x \in \mathbb{R}^n} \; \max_{\lambda, \mu \ge 0} \; \mathcal{L}(x, \lambda, \mu)$$

why? If we choose a feasible point $\tilde{x}$

$h(\tilde{x}) = 0, \; g(\tilde{x}) \le 0$

$$\mathcal{L}(\tilde{x}, \lambda, \mu) = f(\tilde{x}) + \lambda^T \cancel{h(\tilde{x})}^{\,0} + \mu^T g(\tilde{x})$$

$$[\mu_1, \mu_2, \dots, \mu_m] \begin{bmatrix} g_1(\tilde{x}) \\ \vdots \\ g_m(\tilde{x}) \end{bmatrix}$$

case 1: $g_j(\tilde{x}) = 0 \Rightarrow \mu_j$ does nothing

case 2: $g_j(\tilde{x}) < 0 \Rightarrow \mu_j = 0$

$\Rightarrow$ The only way for $\max_{\lambda, \mu \ge 0} \mathcal{L}(x, \lambda, \mu) < +\infty$ is for x to be feasible

---

Now: how do we actually solve these?

1. Penalty method: impose constraints as costs

Quadradtic penalty:

$$\min \; f(x) \quad \text{s.t.} \quad h(x) = 0, \; g(x) \le 0$$

$$\phi_2 = f(x) + \rho \, h(x)^T h(x) + \rho \, g_+(x)^T g_+(x)$$

$\nabla \phi_2$

$\nabla^2 \phi_2$

$$g_+(x) = \max \{ g(x), 0 \}$$

$$g(x) = [1, -1, 0] \Rightarrow g_+(x) = [1, 0, 0]$$

Penalty weight $\rho \to \infty$ risk: numerical conditioning

Alg 1 (Penalty method)

Input: x_0, $\rho_0 > 0$, constraint tol $\tau > 0$

for $k = 0, 1, \dots, k_{max}$

  $x_{k+1} \leftarrow \arg\min \; \phi_2(x; \rho_k)$ (unconst.)

  if $|h_i(x)| < \tau$ and $|g_+^j(x)| < \tau$

    break

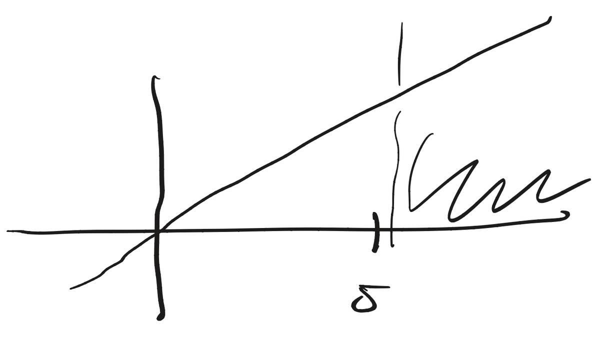
  else:

    choose $\rho_{k+1} > \rho_k$ $(\rho_{k+1} = \alpha \rho_k)$ $\alpha > 1$

Fact: for ϕ_2, $h_i(x_k) \approx \dfrac{\lambda_i^*}{\rho_k}$

e.g. 1

$$\min_x \; x \quad \text{s.t.} \quad x \ge 5$$



$x^* = 5$

$$\phi_2(x) = x + \rho (\max\{5 - x, 0\})^2$$

Augmented Lagrangian (AL)

For an equality constrained problem, define the Augmented Lagrangian

$$\mathcal{L}_A(x, \lambda) = f(x) + \lambda^T h(x) + \frac{\rho}{2} h(x)^T h(x)$$

x_k, λ_k

Alg 1 (Augmented Lagrangian)

Input: x_0, λ_0, $\rho_0 > 0$, $\tau > 0$ tolerance

for $k = 0, 1, \dots, k_{max}$

  $x_{k+1} \leftarrow \arg\min_x \mathcal{L}(x, \lambda_k, \rho_k)$

  $\lambda_{k+1} \leftarrow \lambda_k + \rho_k \, h(x_{k+1})$ (grad ascent $\nabla_\lambda \mathcal{L}_A$)

  if constraints $< \tau$

    return x_{k+1}

  else

    $\rho_{k+1} \ge \rho_k$ $(\rho_{k+1} = \max(\alpha \rho_k, \rho_{max}))$

Note: can apply for ineq, w/ $\tilde{h}(x) = \begin{bmatrix} h(x) \\ g_+(x) \end{bmatrix} = 0$