Lecture 24 (Nov/28)

HW 5 due Thursday

Scribe?

Last time

- CG continued
- Convergence Guarantees
- Nonlinear least squares

Today

- Trust region methods
- Characterization of subproblem
- How about other norms?

Trust region methods

Idea: Instead of fixing a search direction $p_k = B_k^{-1} g_k$, search everywhere near x_k.

Update

$$(\star) \quad s_k = \underset{\text{s.t. } \|s\|_2 \le \Delta_k}{\operatorname{argmin}} \; m_k(s) = f(x_k) + \nabla f(x_k)^T s + \frac{s^T B_k s}{2}$$

$$x_{k+1} = x_k + s_k.$$

In what follows we cover

- A characterization of solutions of $(\star)$
- How about other norms?

- How to solve the subproblem (★)?
- selection of Δ_k and Descent.
- Full Trust Region Method.
- Convergence Guarantees.

By compacteness of $\{s : \|s\|_2 \leq \Delta_k\}$, a minimizer of (★) is well-defined for any B_k (before we needed $B_k \succ 0$).

We obtained indefinite B_k in the past:
- Nonlinear Least Squares (when $\nabla F(x)$ was not full-rank)
- SR1 Quasi-Newton yields indefinite B_k.

Intuitively, if $m_k(s)$ is locally accurate we should obtain descent.

Theorem (♥) (4.1 in Nocedal & Wright)

A vector s^* is a global minimizer of

$$\min \quad f + g^T s + \frac{1}{2} s^T B s$$
$$\text{s.t.} \quad \|s\|_2 \leq \Delta$$

If, and only if, $\|s^*\|_2 \leq \Delta$ and there

Lagrange Multiplier

exists $\lambda \geq 0$ such that

(a) $(B + \lambda I)\, s^* = -g$

(b) $\lambda\, (\Delta - \|s^*\|_2) = 0$ ← Complementary slackness.

(c) $B + \lambda I \succeq 0$

$\dashv$

Remarks

- Necessary and sufficient conditions for nonconvex optimization are rare.

- When $\lambda = 0 \Rightarrow$ (b) allows for $\|s^*\| < \Delta$

  (a) yields $Bs^* + g = 0$

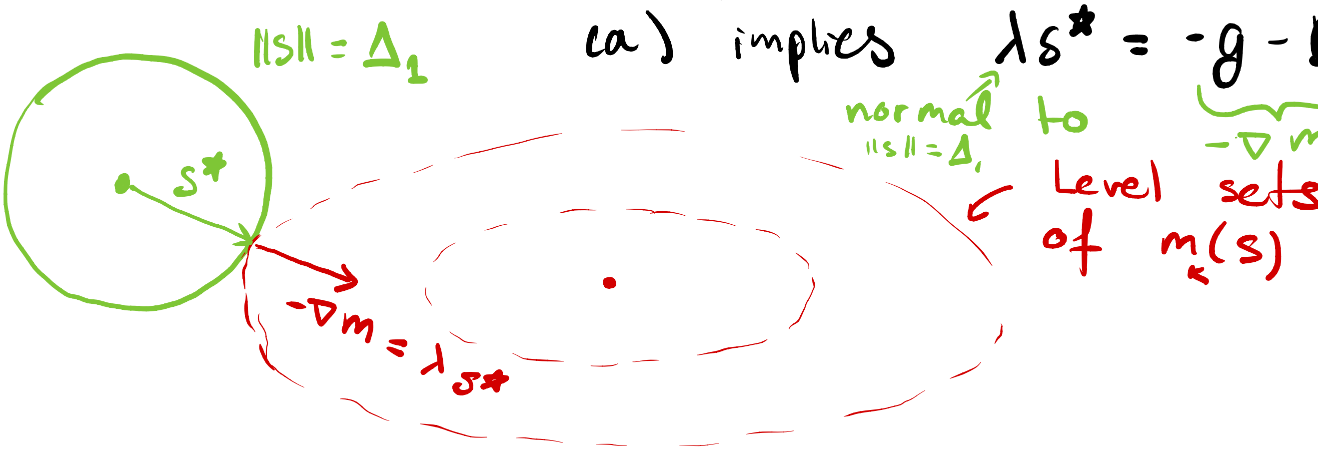
  (1st order unconstrained conditions)

  (c) becomes $B \succeq 0$

  (objective is convex)

- When $\lambda > 0 \Rightarrow$ (b) gives $\|s^*\| = \Delta$.

  (a) implies $\underbrace{\lambda s^*}_{\text{normal to } \|s\| = \Delta_1} = \underbrace{-g - Bs^*}_{-\nabla m}$



▷ Theorem (♡) allows us to algoritmically search for λ.

By (c), $\lambda \geq -\lambda_1$ where the eigenvalues of B are $\lambda_1 < \lambda_2 \leq \dots \leq \lambda_d$

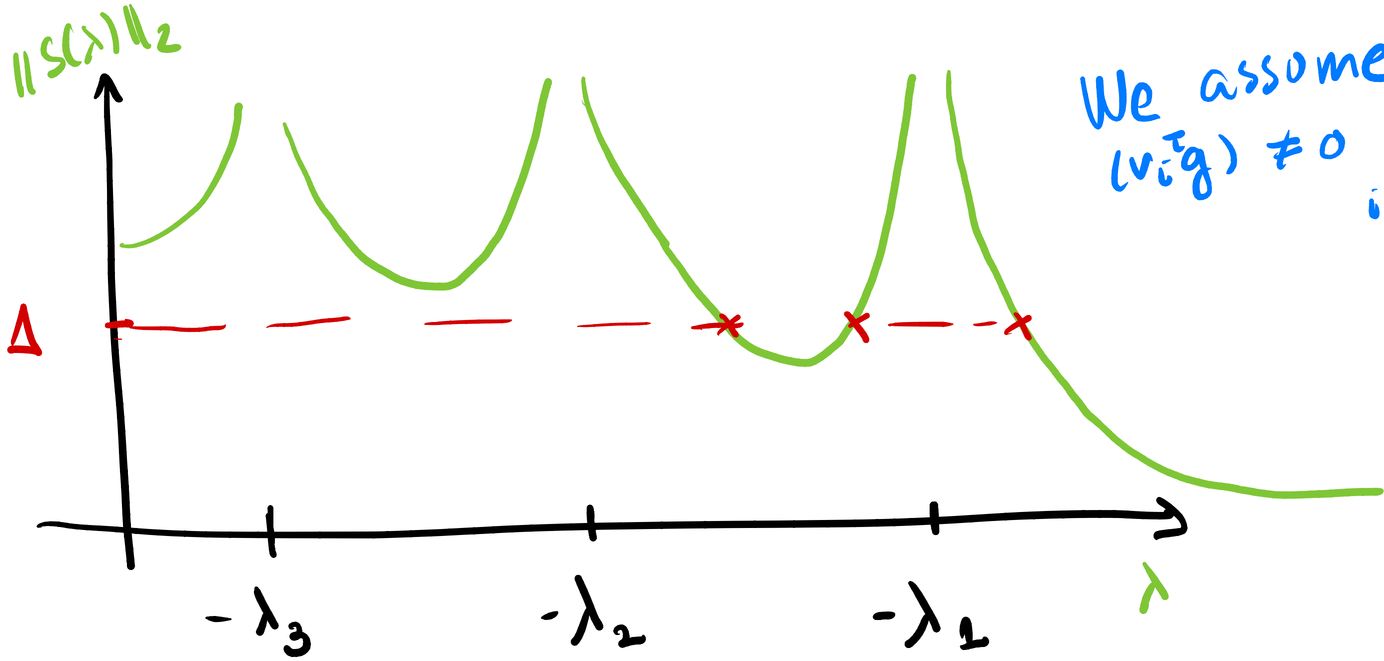
with eigenvales $v_1, \dots, v_d$

Let's search for $\lambda \in (-\lambda_1, \infty)$, define $s(\lambda) = -(B + \lambda I)^{-1} g$. We wish (b) holds, i.e., $\|s(\lambda)\|_2 = \Delta$.

Requires solving linear systems

Note that

$$\|s(\lambda)\|_2^2 = \left\| \sum_{i=1}^{d} \frac{v_i^T g}{\lambda_i + \lambda} v_i \right\|_2^2 = \sum \frac{(v_i^T g)^2}{\lambda_i + \lambda}$$



We assume $(v_i^T g) \neq 0$ for $i = 1, 2, 3$.

We will always have that $\|s(\lambda)\|_2$ is decreasing after $-\lambda_1$. A root-finding method applied to $\|s(\lambda)\|_2 - \Delta$ should yield the unique solution.

Section 4.3. of Nocedal & Wright contains improvements.

Proof of Theorem (♡):

($\Leftarrow$) Let $\lambda \geq 0$ satisfying (a), (b), (c) for some s^*. Consider

$$\hat{m}(s) = f + g^T s + \frac{1}{2} s^T (B + \lambda I) s.$$

By (c), this model is convex.
By (a), s^* minimizes $\hat{m}$ globally.
It is easy to see that

$$\hat{m}(s) = m(s) + \frac{\lambda}{2} \|s\|^2.$$

Thus

$$m(s) \geq m(s^*) + \frac{\lambda}{2}\left(\|s^*\|^2 - \|s\|^2\right)$$

By (b) $\lambda \|s^*\|^2 = \lambda \Delta^2$ check

$$\xrightarrow{} \; = m(s^*) + \underbrace{\frac{\lambda}{2}}_{\geq 0}\underbrace{\left(\Delta^2 - \|s\|^2\right)}_{\geq 0}$$

$$\geq m(s^*).$$

when s feasible

($\Rightarrow$) Suppose s^* is a global minimizer over $\|s\|_2 \leq \Delta$.

If $\|s^*\|_2 < \Delta \Rightarrow s^*$ minimizes $m(s)$ over $\mathbb{R}^d$
and $Bs^* = -g$
$B \succeq 0$.

Check this!

Then (a), (b), (c) hold with $\lambda = 0$.

Thus, we focus on the case $\|s^*\| = \Delta$, which makes (b) hold for free.

We will use a strong duality result that will be covered in Nonlinear 2.

Define

$$L(s,\lambda) = f + g^T s + \tfrac{1}{2} s^T B s + \lambda\left(\|s\|_2^2 - \Delta^2\right)$$

Lets consider two problems

$$p := \inf_{s} \sup_{\lambda \geq 0} L(s,\lambda) \quad \text{and} \quad d := \sup_{\lambda \geq 0} \inf_{s} L(s,\lambda)$$

when a constraint qualification holds (e.g., $\exists\, s$ s.t. $\|s\| < \Delta$) then

$$p = q.$$

Note that

$$\sup_{\lambda \geq 0} L(s, \lambda) = \begin{cases} m(s) & \text{if } \|s\| \leq \Delta \\ +\infty & \text{otherwise.} \end{cases}$$

Similarly

$$\inf_{s} L(s, \lambda) = \begin{cases} \inf \hat{m}(s) - \lambda \Delta^2 & \text{if } (B + \lambda I) \succeq 0 \\ -\infty & \text{otherwise.} \end{cases}$$

Thus there $\exists \lambda \geq 0$ st $B + \lambda I \succeq 0$ (c) and

$$m(s^*) = \inf_{s} \hat{m}(s) - \lambda \Delta^2$$

Since s^* achieves the infimum. Then,

$$\nabla \hat{m}(s^*) = 0 \Rightarrow (B + \lambda I) s^* = -g \quad \text{(a).}$$

$\square$

How about other norms?

The l_2 norm is rather special.
If we use the l_∞ norm, the problem is intractable. Recall

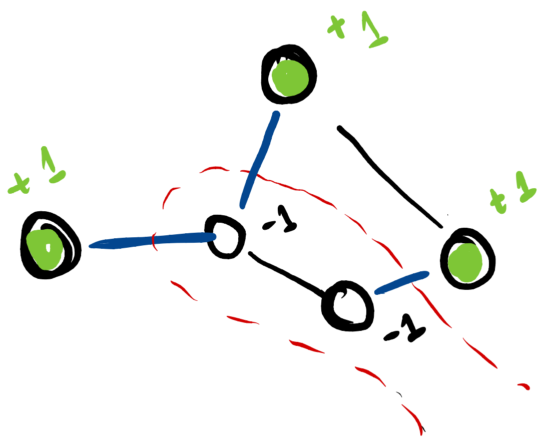
$$\|x\|_\infty := \max_i |x_i|.$$

Theorem. Given a matrix $B \in \mathbb{R}^{d \times d}$ as input. The decision problem that arises from minimizes

$$\min \ x^T B x \quad \text{s.t. } \|x\|_\infty \le \Delta$$

is NP-hard.

Proof: Reduce from MAXCUT, with B the adjacency matrix of the graph:



$$\max_{x \in \{\pm 1\}^n} \sum_{(i,j) \in E} (1 - x_i x_j)$$

$$= \#E - \min \sum_{(i,j)} x_i x_j$$

$$= \#E - \min \ x^T A x$$

$A_{ij} = \mathbb{1}\{(i,j) \in E\}$ $\square$