Trust Region Methods

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

$$(*) \quad s_k = \operatorname{argmin} \; m_k(s) = f(x_k) + \nabla f(x_k)^T s + \frac{1}{2} s^T B_k s$$
$$\text{s.t. } \|s\|_2 \le \Delta_k$$
$$x_{k+1} = x_k + s_k$$

Today
1. How to Solve (*) with Indefinite B_k
2. Other Norms?
3. Selection of Δ_k and Descent

Thursday
4. Full Trust Region Method
5. Convergence Guarantees

1. How to Solve (*) despite Nonconvexities

[Note since we added the constraint $\|s\|_2 \le \Delta$, (*) is well-defined (by compactness) for any B_k (no need for $B_k \succ 0$).]

This shape arose for - Nonlinear Least Squares (when $\nabla F(x)$ was not full rank)

- SR1 Quasi-Newton gave indefinite B_k.

If $m_k(s)$ is locally accurate, then should descend each step.

Theorem (4.1, Nocedal + Wright)

A vector s^* is the global minimizer of

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

Lagrange Multiplier

iff $\|s\|_2 \le \Delta$ and there exist $\lambda \ge 0$ such that

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

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

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

Thoughts/Remarks

- Exact Conditions for Nonconvex OPT are rare.

- When $\lambda = 0$, then (b) allows none tight $\|s\|_2 < \Delta$.

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

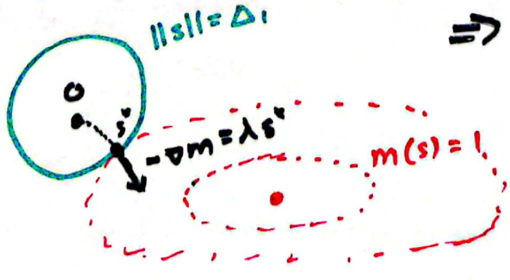
  (first order unconstrained OPT)

  (c) becomes $B \succeq 0$.

  (obj is convex)

- When $\lambda > 0$, $\|s^*\|_2 = \Delta$. Then (a)

$$\Rightarrow \underbrace{\lambda s^*}_{\substack{\text{normal}\\ \text{to constraint}}} = \underbrace{-g - Bs^*}_{\text{negative gradient}}$$



$m_0(s) = ?$

- Algorithmically, we can search for λ.

By (c), $\lambda \geq -\lambda_1$ where the eigenvalues of B are $\lambda_1 \leq \lambda_2 \leq \dots \leq \lambda_d$.
(with eigenvectors $v_1, \dots, v_d$)

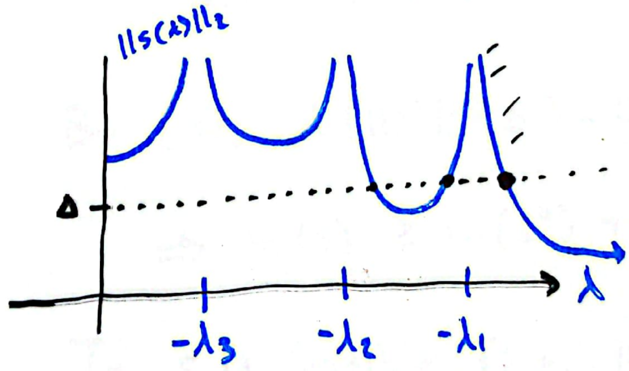
Lets search $\lambda \in (-\lambda_1, \infty)$.

Define $s(\lambda) = -(B + \lambda I)^{-1} g$ (i.e. the solution to (a))

Want (b) to be true, $\|s(\lambda)\|_2 = \Delta$.

Suppose $v_1^T g \neq 0$, $v_2^T g \neq 0$, $v_3^T g \neq 0$

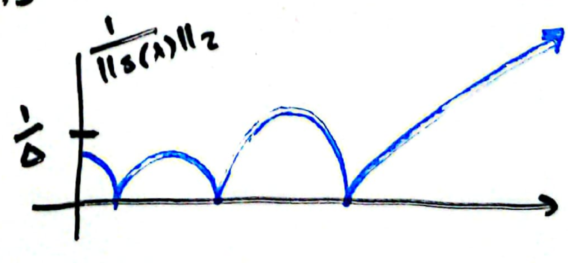
$$\|s(\lambda)\|_2^2 = \left\| \sum_{i=1}^{d} \frac{v_i^T g}{\lambda_i + \lambda} v_i \right\|_2^2 = \sum_{i=1}^{d} \frac{(v_i^T g)^2}{(\lambda_i + \lambda)^2}$$



Missing Squares When Done In Lecture

Nonincreasing after $-\lambda_1$. Root-finding will give the unique solution.

Sec 4.3 of Nocedal+Wright, practical improvements



Proof of Thm 4.1

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

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

By (c), this is convex.

By (a), $\nabla \hat{m}(s^*) = 0$

$\Rightarrow$ s^* globally minimizes $\hat{m}$. (by unconstrained convex optimality)

$\Rightarrow$ For all s,

$$\hat{m}(s) \geq \hat{m}(s^*)$$

Noting $\hat{m}(s) = m(s) + \frac{\lambda}{2} \|s\|_2^2$,

$$m(s) \geq m(s^*) + \frac{\lambda}{2} \|s^*\|_2^2 - \frac{\lambda}{2} \|s\|_2^2.$$

By (b), $\lambda(\|s^*\| - \Delta) = 0$

$$\Rightarrow m(s) \geq m(s^*) + \underbrace{\frac{\lambda}{2}}_{\geq 0} \underbrace{(\Delta^2 - \|s\|_2^2)}_{\geq 0}$$

If s is feasible (i.e. $\|s\|_2 \leq \Delta$),

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

$\Rightarrow$ s^* minimizes m over $\|s\|_2 \leq \Delta$.

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

If $\|s^*\|_2 < \Delta$, then claim s^* minimizes $m(s)$ over $\mathbb{R}^d$

and $Bs^* = -g$

$B \succeq 0$.

Why? Good exercise. Only works for quadratic $m(s)$.

$\Rightarrow$ (a) and (c) with $\lambda = 0$.

($\lambda = 0$ trivially has (b)).

If $\|s^*\|_2 = \Delta$, then (b) is free.

[Need Duality Result to build λ]

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

Note $\sup_{\lambda \ge 0} L(s,\lambda) = \begin{cases} f + g^T s + \frac{1}{2} s^T B s & \text{if } \|s\|_2 \le \Delta \\ +\infty & \text{if } \|s\|_2 > \Delta \end{cases}$

Original Problem $= \min_s \sup_{\lambda \ge 0} L(s,\lambda)$.

attained when $s = s^*$

attained at s^*

$\boxed{\text{Minimax Theorem, need check Constraint Qualification}}$

$= \max_{\lambda \ge 0} \left(\min_s L(s,\lambda) \right)$

attained at s^*.

attained somewhere λ^*

$\Rightarrow s^*$ minimizes globally $L(s, \lambda)$

Necessary Conditions (Second-Order)

$$\nabla_s L(s^*, \lambda^*) = 0 \Leftrightarrow (a)$$

$$\nabla^2_{ss} L(s^*, \lambda^*) \succeq 0 \Leftrightarrow (c). \qquad \square$$

2. Other Norms?

Other norms are often intractable for $(*)$.

Lets examine $\|x\|_\infty = \max_{i=1,\dots,d} |x_i|$.

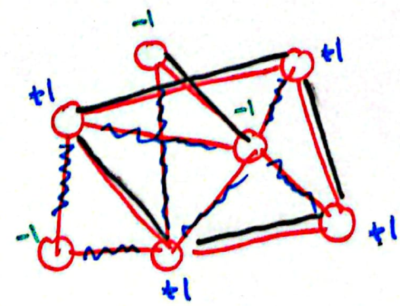
Theorem Given a matrix $B \in \mathbb{R}^{d \times d}$ as input. Solving the minimization

$$\begin{cases} \min & x^T B x = \langle B, xx^T \rangle \cong \langle B, X \rangle_{X \succeq 0} \\ \text{s.t.} & \|x\|_\infty \le \Delta \quad (\Leftrightarrow -\Delta \le x_i \le \Delta) \end{cases}$$

[Goemans Williamson]

is NP-Hard. (Complete if rational)

Proof. Reduce to MAX-CUT, B = Adjacency Matrix of your graph. $\square$



$$x^T B x = \sum_{ij} \begin{cases} 0 & \text{if no edge } ij \\ 1 & \text{if same groups} \\ -1 & \text{if diffent} \end{cases}$$