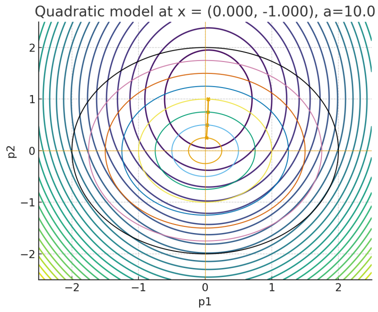
Exercise 3.

1. Let
$$f(x) = 10(x_2 - x_1^2)^2 + (1 - x_1)^2.$$
At $x = (0, -1)$ draw the contour lines of the quadratic model
$$m_k(p) = f_k + g_k^T p + \tfrac{1}{2} p^T B_k p,$$
where B_k is the Hessian of f_k, $g_k = \nabla f(x_k)$ and $f_k = f(x_k)$. Draw the family of solutions of the trust-region subproblem
$$\min_{p \in \mathbb{R}^n} m_k(p) \quad \text{s.t. } \|p\| \le \Delta_k,$$
as the trust region radius Δ varies from 0 to 2. Repeat this at $x = (0, 0.5)$.

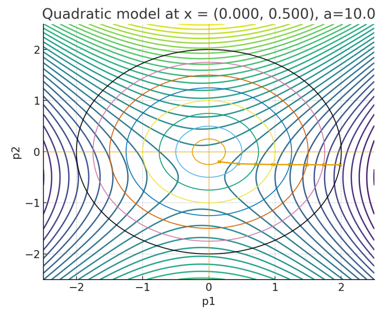
Sol: $\nabla f = \begin{bmatrix} -40 x_1 (x_2 - x_1^2) + 2x_1 - 2 \\ 20 (x_2 - x_1^2) \end{bmatrix}$ $\quad \nabla^2 f = \begin{bmatrix} 120 x_1^2 - 40 x_2 + 2 & -40 x_1 \\ -40 x_1 & 20 \end{bmatrix}$



1) $x_1 = (0, -1)$

$f_1 = 11$ $\quad g_1 = \nabla f_1 = \begin{bmatrix} -2 \\ -20 \end{bmatrix}$. $\quad B_1 = \nabla^2 f_1 = \begin{bmatrix} 42 & 0 \\ 0 & 20 \end{bmatrix} \succ 0$

$p_1 = -B_1^{-1} g_1 = -\frac{1}{840} \begin{bmatrix} 20 & 0 \\ 0 & 42 \end{bmatrix} \begin{bmatrix} -2 \\ -20 \end{bmatrix} = -\frac{1}{840} \begin{bmatrix} -40 \\ -840 \end{bmatrix} = \begin{bmatrix} \frac{1}{21} \\ 1 \end{bmatrix}$



2) $x_2 = (0, \frac{1}{2})$.

$f_2 = \frac{7}{2}$. $\quad g_2 = \nabla f_2 = \begin{bmatrix} -2 \\ 10 \end{bmatrix}$ $\quad B_2 = \begin{bmatrix} -18 & 0 \\ 0 & 20 \end{bmatrix}$ indefinite.

$p_2(\lambda) = -(B_2 + \lambda I)^{-1} g_2 = \begin{bmatrix} \frac{2}{\lambda - 18} \\ -\frac{10}{\lambda + 20} \end{bmatrix}$

$\Rightarrow \left(\frac{2}{\lambda - 18}\right)^2 + \left(\frac{10}{\lambda + 20}\right)^2 = \Delta^2$

the solution p_2^* on the boundary

2. Write a program that implements the dogleg method. Choose B_k to be the exact Hessian. Apply it to solve Rosenbrock's function
$$f(x) = 100(x_2 - x_1^2)^2 + (1 - x_1)^2.$$
Experiment with the update rule for the trust region by changing the constants in the trust-region algorithm below, or by designing your own rules.

Algorithm (Trust Region):

(a) Given $\hat{\Delta} > 0$, $\Delta_0 \in (0, \hat{\Delta})$, and $\eta \in [0, \frac{1}{4})$.

(b) For $k = 0, 1, 2, \ldots$:

i. Obtain p_k by (approximately) solving the subproblem above.

ii. Evaluate ρ_k.

iii. If $\rho_k < \frac{1}{4}$, set $\Delta_{k+1} = \frac{1}{4}\Delta_k$.

iv. Else if $\rho_k > \frac{3}{4}$ and $\|p_k\| = \Delta_k$, set $\Delta_{k+1} = \min(2\Delta_k, \hat{\Delta})$.

v. Else set $\Delta_{k+1} = \Delta_k$.

vi. If $\rho_k > \eta$, set $x_{k+1} = x_k + p_k$, else $x_{k+1} = x_k$.

see the Appendix pls.

3. Theorem 14 (see Lecture Note 4) states that the sequence $\{\|g_k\|\}$ has an accumulation point at zero. Show that if the iterates x_k stay in a bounded set B, then there is a limit point x_∞ of the sequence $\{x_k\}$ such that

$$g(x_\infty) = 0.$$

Pf: $\{x_k\} \subset B$. $\exists \{x_{k_j}\} \subseteq \{x_k\}$ that converges. denote $x_{k_j} \to x_\infty$.

by thm 14, we have $\liminf_k \|g_k\| = 0$ thus. we can find subsequence $\{k_{j_l}\}$ s.t. $\|g_{k_{j_l}}\| \to 0$.

f is Lipschitz cont. $\Rightarrow$ g is Cont. $\Rightarrow$ $g(x_\infty) = \lim_{l\to\infty} g(x_{k_{j_l}}) = 0$.

4. Show that τ_k defined by

$$\tau_k = \begin{cases} 1, & \text{if } g_k^T B_k g_k \leq 0, \\ \min\left(\dfrac{\|g_k\|^3}{\Delta_k g_k^T B_k g_k}, 1\right), & \text{otherwise,} \end{cases}$$

does indeed identify the minimizer of m_k along the direction $-g_k$.

Pf: let $p_k^S = -\dfrac{\Delta_k}{\|g_k\|} g_k$

1) $g_k^T B_k g_k \leq 0$. $m_k(\tau p_k^S)$ decreasing function w.r.t. τ. among $g_k \neq 0$.

thus, $\tau_k = 1$. which is largest value s.t. $\|p_k\| \leq \Delta_k$.

2) $g_k^T B_k g_k > 0$. $m_k(\tau p_k^S)$ convex quadratic w.r.t. τ.

thus. $\tau_k = -\dfrac{g_k^T \cdot p_k^S}{p_k^{S^T} B_k p_k^S} = -\dfrac{\|g_k\|^2 \cdot \frac{\Delta_k}{\|g_k\|}}{g_k^T B_k g_k \cdot \frac{\Delta_k^2}{\|g_k\|^2}} = -\dfrac{\|g_k\|^3}{\Delta_k \cdot g_k^T B_k g_k}$ as a minimizer of convex quadratic

and also needs to satisfy the bound region condition $\tau_k \leq 1$.

thus $\tau_k = \min\left(1, -\dfrac{\|g_k\|^3}{\Delta_k \cdot g_k^T B_k g_k}\right)$.

5. The Cauchy–Schwarz inequality states that for any vectors u and v,

$$|u^T v|^2 \leq (u^T u)(v^T v),$$

with equality only when u and v are parallel. When B is positive definite, use this inequality to show that

$$\gamma \stackrel{\text{def}}{=} \frac{\|g\|^4}{(g^T B g)(g^T B^{-1} g)} \leq 1,$$

with equality only if g, Bg, and $B^{-1}g$ are parallel.

Sol: since $B \succ 0$. $g^T B g > 0$. $g^T B^{-1} g > 0$ for any $g \neq 0$.

also since $B \succ 0$. $B^{1/2}$ exists.

$u = B^{1/2} g$ $v = B^{-1/2} g$.

$\|g\|^4 = (\|g\|^2)^2 = (g^T g)^2 = (u^T v)^2 \leq (u^T u)(v^T v) = (g^T B g)(g^T B^{-1} g) \Rightarrow \gamma \leq 1$.

the equality holds iff. $B^{1/2} g$. $B^{-1/2} g$ parallel

$\Leftrightarrow B^{1/2} g = \lambda B^{-1/2} g \overset{\text{left} \times B^{1/2}}{\Longleftrightarrow} Bg = \lambda g$ $\Leftrightarrow$ g. Bg. $B^{-1} g$ parallel

$\overset{\text{left} \times B^{-1/2}}{\Longleftrightarrow} g = \lambda B^{-1} g$.

6. Show that the following two root-finding updates are equivalent:

$$\lambda^{(\ell+1)} = \lambda^{(\ell)} - \frac{\phi_2(\lambda^{(\ell)})}{\phi_2'(\lambda^{(\ell)})}, \quad \text{(A)}$$

and

$$\lambda^{(\ell+1)} = \lambda^{(\ell)} + \left(\frac{\|p_\ell\|}{\|q_\ell\|}\right)^2 \frac{(\|p_\ell\| - \Delta)}{\Delta}, \quad \text{(B)}$$

using the identities

$$\frac{d}{d\lambda}\left(\|p(\lambda)\|^{-1}\right) = -\tfrac{1}{2}\|p(\lambda)\|^{-3}\frac{d}{d\lambda}\|p(\lambda)\|^2,$$

$$\frac{d}{d\lambda}\|p(\lambda)\|^2 = -2\sum_{j=1}^{n}\frac{(q_j^T g)^2}{(\lambda_j+\lambda)^3},$$

and

$$\|q\|^2 = \sum_{j=1}^{n}\frac{(q_j^T g)^2}{(\lambda_j+\lambda)^3}.$$

Pf: $\phi_2(\lambda) = \frac{1}{\Delta} - \frac{1}{\|p(\lambda)\|}$

$$\phi_2'(\lambda) = \frac{d}{d\lambda}\left(-\frac{1}{\|p(\lambda)\|}\right) = \frac{1}{2}\|p(\lambda)\|^{-3}\frac{d}{d\lambda}\|p(\lambda)\|^2 = -\frac{1}{\|p(\lambda)\|^3}\cdot\sum\frac{(q_j^T g)^2}{(\lambda_j+\lambda)^3} = -\frac{\|q\|^2}{\|p(\lambda)\|^3}$$

$$-\frac{\phi_2(\lambda)}{\phi_2'(\lambda)} = \frac{-\frac{\|p(\lambda)\|-\Delta}{\Delta\|p(\lambda)\|}}{-\|q\|^2\cdot\|p(\lambda)\|^{-3}} = \frac{\|p\|^2}{\|q\|^2}\cdot\frac{\|p\|-\Delta}{\Delta}$$

i.e. $\left(\frac{\|p_\ell\|}{\|q_\ell\|}\right)^2\frac{\|p_\ell\|-\Delta}{\Delta} = -\frac{\phi_2(\lambda^{(\ell)})}{\phi_2'(\lambda^{(\ell)})}$

7. Derive the solution of the two-dimensional subspace minimization problem in the case where B is positive definite.

model: in the 2-dim space $S = \text{span}\{g, B^{-1}g\}$ minimize $m(p) = f + g^T p + \frac{1}{2}p^T B p$. s.t. $\|p\| \leq \Delta$

Sol: let $z_1 = \frac{g}{\|g\|}$ $u := B^{-1}g$ $v = u - z_1^T u z_1$ $z_2 := \frac{v}{\|v\|}$ $(v \neq 0)$.

$\{z_1, z_2\}$. orthonormal basis for S.

$\forall p \in S$. $\exists!\, y \in \mathbb{R}^2$ s.t. $p = [z_1\ z_2]y$ and $\|p\| = \|[z_1, z_2]y\| = \|y\|$

denote $g_S := [z_1\ z_2]g$. $H_S := [z_1\ z_2]^T B [z_1, z_2]$. $H_S \succ 0$.

thus the original problem equivalent to $\min_{\|y\|\leq\Delta} \widehat{m}(y) := f + g_S^T y + \frac{1}{2}y^T H_S y$.

i) $\|y\| < \Delta$.

$\nabla\widehat{m}(y) = 0 \Rightarrow y^* := -H_S^{-1}g_S$ $\|y^*\| \leq \Delta$.

$\Rightarrow p^* = [z_1\ z_2]\cdot(-H_S^{-1}g_S)$

ii) $\|y\| = \Delta$.

$\exists\lambda$. $g_S + H_S y + \lambda y = 0 \Rightarrow y^*$ s.t. $(H_S + \lambda I)y = -g_S$. for some $\lambda \in \mathbb{R}$.

8. Show that if B is any symmetric matrix, then there exists $\lambda \geq 0$ such that $B + \lambda I$ is positive definite.

Pf: let B has eigenvalue $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_k$. set $\lambda = -|\lambda_k| + \delta$.

$\forall v \in \mathbb{R}^n$ $(B+\lambda I)v = Bv + \lambda I v = \lambda_i v + \lambda v = (\lambda_i + \lambda)v$. for every $i \in [1:k]$.

then the eigenvalue of B are $\{\lambda_1 + \lambda, \ldots, \lambda_k + \lambda\}$.

let $\{v_1, \ldots v_n\}$ be the set of eigenvectors of $B + \lambda I$. mutually orthogonal.

$\forall v \in \mathbb{R}^n \setminus \{0\}\ \exists\{c_1, \ldots c_n\}$. $v = \sum c_i v_i$.

$v^T(B+\lambda I)v = (\sum c_i v_i)^T(B + \lambda I)(\sum c_i v_i) = \sum_{i=1}^{n} c_i^2(\lambda_i + \lambda).\|v_i\|^2 > 0$

thus $B + \lambda I \succ 0$.

9. Verify that the definitions

$$p_k^S = -\frac{\Delta_k}{\|D^{-1}g_k\|}D^{-2}g_k,$$

and

$$\tau_k = \begin{cases} 1, & \text{if } g_k^T D^{-2}B_k D^{-2}g_k \le 0, \\ \min\left(\dfrac{\|D^{-1}g_k\|^3}{\Delta_k g_k^T D^{-2}B_k D^{-2}g_k}, 1\right), & \text{otherwise,} \end{cases}$$

are valid for the Cauchy point in the case of an elliptical trust region, where D is a diagonal scaling matrix with positive diagonal elements.

the scaled TR-subproblem.

$$\min_{p\in\mathbb{R}^n} m_k(p) = f_k + g_k^T p + \frac{1}{2}p^T B_k p. \quad \text{s.t. } \|Dp\| \le \Delta_k.$$

1) $g_k^T D^{-2} B_k D^{-2} g_k \le 0$ $m_k(\tau p_k^S)$ decreasing function w.r.t. τ. among $g_k \ne 0$.

thus, $\tau_k = 1$. which is largest value s.t. $\|\tau D p_k^S\| \le \Delta_k$.

2). $g_k^T D^{-2} B_k D^{-2} g_k > 0$ D p.d. and diag. $(D^{-1})^T = D^{-1}$

$$\tau_k = -\frac{g_k^T \cdot p_k^S}{p_k^{S T} B_k p_k^S} = -\frac{g_k^T \cdot -\frac{\Delta_k}{\|D^{-1}g_k\|} D^{-2} g_k}{\frac{\Delta_k^2}{\|D^{-1}g_k\|^2} g_k^T D^{-2} B_k D^{-2} g_k} = \frac{(D^{-1}g_k)^T(D^{-1}g_k)\cdot\|D^{-1}g_k\|}{\Delta_k \cdot g_k^T D^{-2} B_k D^{-2} g_k}$$

$$= \frac{\|D^{-1}g_k\|^3}{\Delta_k \cdot g_k^T D^{-2} B_k D^{-2} g_k}$$ as a minimizer of convex quadratic

and also needs to satisfy the bound region condition $\tau_k \le 1$.

thus $\tau_k = \min\left(1, \frac{\|D^{-1}g_k\|^3}{\Delta_k \cdot g_k^T D^{-2} B_k D^{-2} g_k}\right)$.

10. Consider the trust-region subproblem

$$\min_{p\in\mathbb{R}^n} m_k(p) = f_k + g_k^T p + \tfrac{1}{2}p^T B_k p \quad \text{s.t. } \|p\| \le \Delta_k.$$

Suppose B_k is positive definite. Prove that if the unconstrained minimizer $p_B = -B_k^{-1}g_k$ satisfies $\|p_B\| \le \Delta_k$, then p_B is also the solution of the trust-region subproblem. Otherwise, the solution lies on the boundary $\|p\| = \Delta_k$.

Pf: $B_k \succ 0$. we set $\lambda = 0$ in Moré - Sorensen thm

i/ $\|p_B\| \le \Delta_k$. p_B is feasible

$m_k(p)$ convex quadratic. minimizer $p_B = -B_k^{-1} g_k$

by Moré - Sorensen thm we have p_B is the solution

ii/ $\|p_B\| > \Delta_k$. assume $\exists\, p^*$. the solution s.t. $\|p^*\| < \Delta_k$.

by the condition $(B + \lambda I)p^* = -g$ $\Rightarrow$ $Bp^* + g = 0$ $\Rightarrow$ $p^* = p_B$. contradicts with $\|p_B\| > \Delta_k$

thus $\|p^*\| = \Delta_k$ m_k convex quadratic