Convex Analysis with Applications
UBC Math 604 Lecture Notes by Philip D. Loewen

In trust region methods, we minimize a quadratic model function $M = M(p)$ over the set of all $p \in \mathbb{R}^n$ satisfying a constraint

$$g(p) \overset{\text{def}}{=} \frac{1}{2} \left( ||p||^2 - \Delta \right) \leq 0.$$

(Here $\Delta > 0$ is given.) In cases where $M$ is convex, there is a nice theory for this problem; the theory has much more general applicability too.

A. Convex Sets and Functions

**Definition.** A subset $S$ of a real vector space $X$ is called **convex** when, for every pair of distinct points $x_0, x_1$ in $S$, one has

$$x_t := (1-t)x_0 + tx_1 \in S \quad \forall t \in (0, 1).$$

We call $S$ **strictly convex** if in fact one has $x_t \in \text{int } S$ above.

**Definition.** Let $X$ be a real vector space containing a convex subset $S$. A function $f: S \to \mathbb{R}$ is called **convex** when, for every pair of distinct points $x_0, x_1$ in $S$, one has

$$f((1-t)x_0 + tx_1) \leq (1-t)f(x_0) + tf(x_1) \quad \forall t \in (0, 1).$$

If this statement holds with strict inequality, then $f$ is called **strictly convex** on $S$.

Geometrically, a convex function is one whose chords lie on or above its graph; a strictly convex function is one whose chords lie strictly above the graph except at their endpoints.

**Example.** Given a row vector $g$ of length $n$, and an $n \times n$ matrix $H = H^T$ take $X = \mathbb{R}^n$ and $S = X$. Then $f(x) \overset{\text{def}}{=} gx + \frac{1}{2}x^THx$ is convex iff $H \succeq 0$ and strictly convex iff $H > 0$.

**Proof.** Pick $x_0, x_1 \in \mathbb{R}^n$ distinct. Let $v = x_1 - x_0$, and consider, for $t \in (0, 1),

$$x_t = (1-t)x_0 + tx_1 = x_0 + t(x_1 - x_0) = x_0 + tv.$$ We have

$$[(1-t)f(x_0) + tf(x_1)] - f(x_t)$$

$$= (1-t) \left[ gx_0 + \frac{1}{2}x_0^THx_0 \right] + t \left[ gx_1 + \frac{1}{2}x_1^THx_1 \right] - \left[ gx_t + \frac{1}{2}x_t^THx_t \right]$$

$$= \frac{1}{2}(1-t)x_0^THx_0 + \frac{1}{2}tx_1^THx_1 - \frac{1}{2} \left[ (1-t)^2x_0^THx_0 + 2t(1-t)x_0^THx_1 + t^2x_1^THx_1 \right]$$

$$= \frac{1}{2}[(1-t) - (1-t)^2]x_0^THx_0 + \frac{1}{2}[t - t^2]x_1^THx_1 - t(1-t)x_0^THx_1$$

$$= \frac{1}{2}t(1-t)(x_1 - x_0)^TH(x_1 - x_0) = \frac{1}{2}tv^THv.$$ Now $v \neq 0$ by assumption, so if $H > 0$ then the right side is positive whenever $t \in (0, 1)$. If we know only $H \succeq 0$, then at least the right side is nonnegative in this interval.
Notice that the definition of convexity makes no reference to differentiability, and indeed many convex functions fail to be differentiable. A simple example of this is $f_1(x) = |x|$ on $X = \mathbb{R}$—note that $f_1(x) = \int_0^x \text{sgn}(t) \, dt$; a more complicated example is $f_2(x) := \int_0^x \lfloor t \rfloor \, dt$. The graph of $f_2$ is shown below: it has corners at every integer value of $x$.

```
\begin{align*}
f_2(x) &:= \int_0^x \lfloor t \rfloor \, dt.
\end{align*}
```

**Operations Preserving Convexity.** Here are several ways to combine functions that respect the property of convexity. In each of them, $X$ is a real vector space with convex subset $D$. (Confirming these statements is an easy exercise: please try it!)

(a) Positive Linear Combinations: If the functions $f_1, f_2, \ldots, f_n$ are convex on $D$, and $c_1, c_2, \ldots, c_n$ are *positive* real numbers, then the sum function

$$f(x) := c_0 + c_1 f_1(x) + c_2 f_2(x) + \cdots + c_n f_n(x)$$

is well-defined and convex on the set $D$. Moreover, if any one of the functions $f_k$ happens to be strictly convex, then the sum function $f$ will be strictly convex too.

(b) Restriction to Convex Subsets: If $f$ is convex on $D$ and $S$ is a convex subset of $D$, then the restriction of $f$ to $S$ is convex on $S$. (An important particular case arises when $D = X$ is the whole space and $S$ is some affine subspace of $X$: a function convex on $X$ is automatically convex on $S$.)

(c) Pointwise Maxima: If $f_1, f_2, \ldots, f_n$ are convex on $D$, then so is their maximum function

$$f(x) := \max \{ f_1, f_2, \ldots, f_n \}.$$  

(A similar result holds for the maximum of an infinite number of convex functions, under suitable hypotheses.)
B. Convexity and its Characterizations

The First-Derivative Test.

**Theorem (Subgradient Inequality).** Let $X$ be a real vector space containing a convex subset $S$. Let the function $f : X \to \mathbb{R}$ be Gâteaux differentiable at every point of $S$.

(a) $f$ is convex on $S$ if and only if $\nabla f(x)(y - x) \leq f(y) - f(x)$ for any distinct $x$ and $y$ in $S$.

(b) $f$ is strictly convex on $S$ if and only if $\nabla f(x)(y - x) < f(y) - f(x)$ for any distinct $x$ and $y$ in $S$.

**Proof.** (a)(b)$\implies$ Let $x_0$ and $x_1$ be distinct points of $S$, and let $t \in (0, 1)$ be given. Define $x_t := (1 - t)x_0 + tx_1$. In statement (b), the assumption is equivalent to $f(x) < f(y) + \nabla f(x)(x - y)$ for any $x \neq y$: choosing $y = x_t$ here, and then applying the inequality to both $x = x_0$ and $x = x_1$ in turn, we get

\[
\begin{align*}
  f(x_t) &< f(x_0) + \nabla f(x_t)(x_t - x_0), \\
  f(x_t) &< f(x_1) + \nabla f(x_t)(x_t - x_1).
\end{align*}
\]

Multiplying (†) by $(1 - t)$ and (‡) by $t$ and adding the resulting inequalities produces

\[
f(x_t) < tf(x_1) + (1 - t)f(x_0) + \nabla f(x_t)(t(x_t - x_1) + (1 - t)(x_t - x_0)).
\]

A simple calculation reveals that the operand of $\nabla f(x_t)$ appearing here is the zero vector, so this reduces to the inequality defining the strict convexity of $f$ on $S$. In statement (a), the assumed inequality is non-strict, but exactly the same steps lead to a non-strict version of conclusion ($\ast$). This confirms the ordinary convexity of $f$ on $S$.

(a)$\implies$ Suppose $f$ is convex on $S$. Let two distinct points $x$ and $y$ in $S$ be given. Then for any $t \in (0, 1)$, the definition of convexity gives

\[
f(x + t(y - x)) = f((1 - t)x + ty) \\
\leq (1 - t)f(x) + tf(y) = f(x) + t(f(y) - f(x)).
\]

Rearranging this gives

\[
\frac{f(x + t(y - x)) - f(x)}{t} \leq f(y) - f(x).
\]

In the limit as $t \to 0^+$, the left side converges to the directional derivative $f'[x; y - x]$, which equals $\nabla f(x)(y - x)$ by the definition of the Gâteaux derivative. Thus we have

\[
\nabla f(x)(y - x) \leq f(y) - f(x).
\]
This argument applies to any distinct points \( x \) and \( y \) in \( S \), so the result follows.

(b)(\( \Rightarrow \)) If \( f \) is known to be strictly convex on \( S \), then we certainly obtain (**) . It remains only to show that equality cannot hold. Indeed, if \( x_0 \) and \( x_1 \) are distinct points of \( S \) where one has

\[
f(x_1) = f(x_0) + \nabla f(x_0)(x_1 - x_0),
\]

the definition of convexity implies that for all \( t \) in \((0,1)\),

\[
f(x_t) \leq (1 - t)f(x_0) + tf(x_1)
\]

\[
= (1 - t)f(x_0) + t[f(x_0) + \nabla f(x_0)(x_1 - x_0)]
\]

\[
= f(x_0) + t\nabla f(x_0)(x_1 - x_0) = f(x_0) + \nabla f(x_0)(x_t - x_0)
\]

\[
\leq f(x_t).
\]

(The last inequality is a consequence of (**).) This chain of inequalities reveals that

\[
f(x_t) = (1 - t)f(x_0) + tf(x_1) \quad \forall t \in (0,1),
\]

which cannot happen for a strictly convex function \( f \).

The inequality at the heart of the previous Theorem, namely

\[
\nabla f(x)(y - x) \leq f(y) - f(x) \quad \forall y \in X,
\]

is called the subgradient inequality. Geometrically, it says that the tangent hyperplane to the graph of \( f \) at the point \((x, f(x))\) lies on or below the graph of \( f \) at every point \( y \) in \( X \). (Write \( f(y) \geq f(x) + \nabla f(x)(y - x) \) to see this.) The name also helps one remember the direction of the inequality: the “gradient” \( \nabla f(x) \) is always on the small side. It is worth noting that the subgradient inequality written above follows from the convexity of \( f \) using the differentiability property only at the point \( x \)—see the proof of “(a)\( \Rightarrow \)” above.

The Second-Derivative Test. We will never need second derivatives in general vector spaces, so we focus on the finite-dimensional case \((X = \mathbb{R}^n)\) in the next result. Here we use the familiar notation \( \nabla f(x) \) instead of \( \nabla f \), and write \( \nabla^2 f(x) \) for the usual Hessian matrix of mixed partial derivatives of \( f \) at the point \( x \):

\[
[\nabla^2 f(x)]_{ij} = \frac{\partial^2 f}{\partial x_i \partial x_j}.
\]

We assume \( f \in C^2(\mathbb{R}^n) \), so the order of mixed differentiation does not matter and the Hessian matrix is symmetric. The shorthand \( \nabla^2 f(x) > 0 \) means that this matrix is positive definite, i.e.,

\[
v^T \nabla^2 f(x) v > 0 \quad \forall v \in \mathbb{R}^n, \ v \neq 0.
\]  

(B.1)

(A similar understanding governs the notation \( \nabla^2 f(x) \geq 0 \).)
Theorem. Let \( f : \mathbb{R}^n \to \mathbb{R} \) be of class \( C^2 \).

(a) \( f \) is convex on \( \mathbb{R}^n \) if and only if \( \nabla^2 f(x) \geq 0 \) for all \( x \) in \( \mathbb{R}^n \);

(b) If \( \nabla^2 f(x) > 0 \) for all \( x \) in \( \mathbb{R}^n \), then \( f \) is strictly convex on \( \mathbb{R}^n \).

Proof. (b) Pick any two distinct points \( x \) and \( y \) in \( \mathbb{R}^n \). Taylor’s theorem says that there is some point \( \xi \) on the line segment joining \( x \) to \( y \) for which

\[
 f(y) = f(x) + \nabla f(x)(y - x) + \frac{1}{2}(y - x)^T \nabla^2 f(\xi)(y - x).
\]

The inequality here holds because \( \nabla^2 f(\xi) > 0 \) by assumption. Since \( x \) and \( y \) in \( X \) are arbitrary, this establishes the hypothesis of the previous Theorem (part (b)): the strict convexity of \( f \) follows.

(a) \((\Leftarrow)\) Just rewrite the proof of (b) with ‘\( \geq \)’ in place of ‘\( > \)’.

\((\Rightarrow)\) Let us prove first that the operator \( \nabla f \) is monotone, i.e., that

\[
 (\nabla f(y) - \nabla f(x))(y - x) \geq 0 \quad \forall x, y \in \mathbb{R}^n. \tag{*}
\]

To see this, just write down the subgradient inequality twice and add:

\[
\begin{align*}
 f(y) - f(x) &\geq \nabla f(x)(y - x) \\
 f(x) - f(y) &\geq \nabla f(y)(x - y)
\end{align*}
\]

\[
0 \geq (\nabla f(x) - \nabla f(y))(y - x)
\]

Rearranging the sum produces \((*)\).

Now fix any \( x \) and \( v \) in \( \mathbb{R}^n \), and put \( y = x + tv \) into \((*)\), assuming \( t > 0 \). It follows that

\[
 (\nabla f(x + tv) - \nabla f(x))(v) \geq 0 \quad \forall t > 0.
\]

Take \( g(t) := \nabla f(x + tv)(v) \): then the previous inequality implies, upon division by \( t > 0 \), that

\[
\frac{g(t) - g(0)}{t} \geq 0 \quad \forall t > 0.
\]

Now \( f \) is \( C^2 \), so \( g \) is \( C^1 \): hence the limit of the left-hand side as \( t \to 0^+ \) exists and equals \( g'(0) \). We deduce that \( g'(0) \geq 0 \), i.e.,

\[
v^T \nabla^2 f(x)v \geq 0.
\]

Since \( v \in \mathbb{R}^n \) was arbitrary, this shows that \( \nabla^2 f(x) \geq 0 \), as required. // // //

The second-derivative test, together with the simple convexity-preserving operations mentioned above, is an easy way to confirm (or refute) the convexity of a given function. For example, when \( X = \mathbb{R}^n \), the quadratic function

\[
 f(x) := x^THx + gx + c
\]

built around symmetric \( n \times n \) matrix \( H \), a row vector \( g \), and a constant \( c \), will be convex if and only if \( H \geq 0 \) and strictly convex if and only if \( H > 0 \).
Exercises. (1) Consider the $2 \times 2$ matrix $H = \begin{bmatrix} a & b \\ b & d \end{bmatrix}$. Prove that

$$H \geq 0 \iff \begin{bmatrix} a \geq 0, \\ d \geq 0, \\ ad - b^2 \geq 0 \end{bmatrix}. \quad \text{(B.2)}$$

Then show that (B.2) remains true if all strict inequalities (">") are replaced by nonstrict ones ("≥").

(2) Consider the $2n \times 2n$ matrix $H = \begin{bmatrix} A & B \\ B^T & D \end{bmatrix}$, in which $A = A^T$, $B$, and $D = D^T$ are $n \times n$ matrices. Show that

$$H > 0 \iff \begin{bmatrix} A > 0, \\ D > 0, \\ A - BD^{-1}B^T > 0 \end{bmatrix}.$$  

Here the matrix inequalities are interpreted as in (B.1). Hints: (i) $D > 0$ implies the existence of $D^{-1}$; (ii) $D = D^T$ implies $(D^{-1})^T = (D^T)^{-1}$; and (iii) if $D$ is invertible then

$$H = \begin{bmatrix} I & BD^{-1} \\ 0 & I \end{bmatrix} \begin{bmatrix} A - BD^{-1}B^T & 0 \\ 0 & D \end{bmatrix} \begin{bmatrix} I & 0 \\ D^{-1}B^T & I \end{bmatrix}. \quad \text{C. Convexity as a Sufficient Condition—General Theory}

Proposition. Let $S$ be an affine subset of a real vector space $X$. Suppose that $f: S \to \mathbb{R}$ is convex on $S$. If $\hat{x}$ is a point of $S$ where $f$ is Gâteaux differentiable, then the following statements are equivalent:

(a) $\hat{x}$ minimizes $f$ over $S$, i.e., $f(\hat{x}) \leq f(x)$ for all $x$ in $S$;

(b) $\hat{x}$ is a critical point for $f$ relative to $S$, i.e., $\nabla f(\hat{x})(h) = 0$ for all $h$ in the subspace $V = S - \hat{x}$.

Proof. ($a \Rightarrow b$) Proved much earlier, without using convexity.

($b \Rightarrow a$) Note that for any point $x$ of $S$, the difference $h = x - \hat{x}$ lies in the subspace $V$ of the statement. Hence the subgradient inequality, together with condition (b), gives

$$f(x) \geq f(\hat{x}) + \nabla f(\hat{x})(x - \hat{x}) = f(\hat{x}) + 0 \quad \forall x \in S.$$

Corollary. If $f: S \to \mathbb{R}$ is strictly convex on some affine subset $S$ of a real vector space $X$, and $\hat{x}$ is a critical point for $f$ relative to $S$, then $\hat{x}$ minimizes $f$ over $S$ uniquely, i.e., $f(\hat{x}) < f(x)$ for all $x \in S$, $x \neq \hat{x}$.

D. Separation Theorem

Separation. Let $U, V$ be convex subsets of $\mathbb{R}^n$. If the intersection $U \cap V$ contains at most one point, then there exists $\alpha \neq 0$ in $\mathbb{R}^n$ such that

$$\alpha^T u \geq \alpha^T v \quad \forall u \in U, \ v \in V.$$  

More generally, the existence of $\alpha \neq 0$ is assured whenever $(\text{int } U) \cap (\text{int } V) = \emptyset$. 
E. Inequality-Constrained Convex Programming

**Problem Statement.** Differentiable convex functions \( F, G : \mathbb{R}^n \to \mathbb{R} \) are given. We want to
\[
\min \{ F(x) : G(x) \leq 0 \}.
\]

**Sufficient Conditions.** Given \( \hat{x} \in \mathbb{R}^n \), the following conditions guarantee that \( \hat{x} \) solves the problem above:

\[
\begin{align*}
\text{(i) } & \lambda G(\hat{x}) = 0, \\
\text{(ii) } & \text{the function } x \mapsto F(x) + \lambda G(x) \text{ has gradient } 0 \text{ at } \hat{x}.
\end{align*}
\]

Note that the function described above would be convex, so having gradient 0 at \( \hat{x} \) is the same as having a global min at \( \hat{x} \).

**Proof.** Note that since \( \lambda \geq 0 \) and both \( f, g \) are convex, every critical point for \( f + \lambda g \) is a global minimizer:
\[
F(\hat{x}) + \lambda G(\hat{x}) \leq F(x) + \lambda G(x) \quad \forall x \in \mathbb{R}^n.
\]
Now \( G(\hat{x}) \leq 0 \) by assumption, and for any \( x \in \mathbb{R}^n \) satisfying the constraint \( G(x) \leq 0 \),
\[
F(\hat{x}) \leq F(x) + \lambda G(x) \leq F(x).
\]
This completes the proof. /////

**Necessary Conditions.** Suppose \( \hat{x} \in \mathbb{R}^n \) solves the problem above. Then one of these two statements hold:

(i) \( G(\hat{x}) = 0 \) and \( \nabla G(\hat{x}) = 0 \). [ABNORMAL--constraint is degenerate.]

(ii) Conclusion \((*)\) above holds for \( \hat{x} \). [NORMAL]

**Proof.** If \( G(\hat{x}) < 0 \), then \( G(x) < 0 \) for all \( x \) near \( \hat{x} \), so \( \hat{x} \) is an unconstrained local minimum for \( f \). Hence \( \nabla F(\hat{x}) = 0 \). Case (ii) holds, because \( \lambda = 0 \) works in \((*)\).

So assume that \( G(\hat{x}) = 0 \). [Note that conclusion (i) in \((*)\) is then automatic.]
Define
\[
U = \{(F(\hat{x}) - \beta, G(\hat{x}) - \gamma) : \beta > 0, \gamma \geq 0\},
\]
\[
V = \{(F(x) + r, G(x) + s) : x \in \mathbb{R}^n, r, s \geq 0\}.
\]
These are convex sets in \( \mathbb{R}^2 \). [Practice: If \( v_0, v_1 \in V \), say \( v_i = (F(x_i) + r_i, G(x_i) + s_i) \), show \( v_t = (1-t)v_0 + tv_1 \in V \) using convexity inequality on \( f, g \).] Note that \( U \cap V = \emptyset \): if \( (p, q) \in U \cap V \), then there must be some specific \( x \in \mathbb{R}^n \) and \( \alpha > 0, \beta \geq 0, r \geq 0, s \geq 0 \) such that
\[
p = F(\hat{x}) - \beta = F(x) + r, \quad \text{i.e.,} \quad F(x) = F(\hat{x}) - \beta - r < F(\hat{x}),
\]
\[
q = G(\hat{x}) - \gamma = G(x) + s, \quad \text{i.e.,} \quad G(x) = G(\hat{x}) - \gamma - s \leq 0.
\]
The existence of such an \( x \) contradicts the optimality of \( \hat{x} \).

So pick \( \alpha = (\alpha_0, \alpha_1) \neq (0,0) \) such that \( \alpha^T u \leq \alpha^T v \) for all \( u \in U, v \in V \). Expand \( \alpha^T(v - u) \geq 0 \):

\[
0 \leq \alpha_0 [F(x) + r - F(\hat{x}) + \beta] + \alpha_1 [G(x) + s - G(\hat{x}) + \gamma] \\
\forall x \in \mathbb{R}^n, \beta > 0, \gamma, r, s \geq 0. \quad (**)
\]

Pick \( x = \hat{x}, r = 0 = s \):

\[
0 \leq \alpha_0 \beta + \alpha_1 \gamma \quad \forall \beta > 0, \gamma \geq 0.
\]

This forces \( \alpha_0 \geq 0 \) and \( \alpha_1 \geq 0 \).

If \( \alpha_0 = 0 \), then \( \alpha_1 > 0 \) (since vector \( \alpha \neq 0 \)). Take \( s = 0 = \gamma \) in (**) to get

\[
0 \leq G(x) - G(\hat{x}) \quad \forall x \in \mathbb{R}^n.
\]

This is the abnormal case.

Assume \( \alpha_0 > 0 \). Define \( \lambda = \alpha_1/\alpha_0 \geq 0 \). Taking \( \gamma = 0 = r = s \) in (**) and rearranging gives

\[
F(\hat{x}) + \lambda G(\hat{x}) \leq F(x) + \lambda G(x) + \beta \quad \forall \beta > 0.
\]

This implies the same inequality even when \( \beta = 0 \), and shows that \( x \mapsto f + \lambda g \) has a global minimum at \( \hat{x} \). Consequently its gradient is 0: that’s condition (ii) in (*).

To illustrate this proof, here is a sketch of the set \( V \) for the case of \( F(x) = (x - 2)^2 + (y - 2)^2 \) and \( G(x) = x^2 + y^2 - 6 \). The pairs \( (F(x), G(x)) \) are shaded in gray; points included only because of the additional \( (r, s) \) in the definition are shown in black.

![Diagram of the set V]

**Sensitivity.** You can see from picture that, in the presence of sufficient smoothness, the multiplier \( \lambda \) has the following interpretation:

\[
\lambda = -V'(0), \quad \text{where} \quad V(z) = \min \{F(x) : G(x) = z\}.
\]
Here $V$ is the value function associated with our problem. Another proof:

\[
\begin{align*}
V(G(x)) &\leq F(x) \quad \forall x, \\
V(G(\hat{x})) = F(\hat{x}),
\end{align*}
\]

\[\implies f - V \circ g \text{ min at } \hat{x}\]

\[\implies \nabla F(\hat{x}) - V'(0) \nabla G(\hat{x}) = 0.\]

**The Nonconvex Case.** Similar results hold when $f, g$ are nonconvex, but the proofs are much harder. (All results are expressed in terms of critical points rather than global minima.) Detailed discussion postponed.

**F. Application to the Trust-Region Subproblem**

Recall our problem, with $\Delta > 0$ given in advance:

\[
\begin{align*}
\text{minimize} & \quad M(p) \overset{\text{def}}{=} \frac{1}{2}p^T Hp + gp + c \\
\text{subject to} & \quad \|p\| \leq \Delta.
\end{align*}
\]

This has the form above, with a quadratic function $M$ and a constraint $G(p) \leq 0$ where $G(p) = \frac{1}{2} \left(\|p\|^2 - \Delta^2\right)$. Note $\nabla M(p) = g + p^T H$, $\nabla G(p) = p^T$.

**Theorem.** A vector $\hat{p} \in \mathbb{B}[0; \Delta]$ solves the problem above if and only if there exists $\lambda \geq 0$ satisfying all of

(i) \quad $\lambda (\|p\| - \Delta) = 0$,

(ii) \quad $(H + \lambda I)\hat{p} = -g^T$,

(iii) \quad $H + \lambda I \geq 0$.

**Proof.** ($\Leftarrow$) Suppose some $\hat{p}$ obeys (i)–(iii). Then (iii) implies that the quadratic function $M + \lambda G$ is convex, and (i)–(ii) show that $\hat{p}$ is a critical point for this function with $\lambda G(\hat{p}) = 0$. The sufficient conditions above imply that $\hat{p}$ gives the constrained minimum.

($\Rightarrow$) Now suppose $\hat{p} \in \mathbb{B}[0; \Delta]$ gives the minimum. If $\|\hat{p}\| < \Delta$, then conditions (i)–(iii) hold for $\lambda = 0$. So assume $\|\hat{p}\| = \Delta$. Apply the Necessary Conditions above. The abnormal case is impossible—it requires both $G(\hat{p}) = 0$ (so $\|\hat{p}\| = \Delta > 0$) and $\nabla G(\hat{p}) = 0$ (so $\hat{p} = 0$), and these are incompatible. Hence there must be some $\lambda \geq 0$ such that $\lambda G(\hat{p}) = 0$ and

\[0 = \nabla (M + \lambda g)(\hat{p}) = g + p^T H + \lambda p^T, \quad \text{i.e.,} \quad (H + \lambda I)p = -g^T.\]

This proves (i)–(ii). To get (iii), use (ii) to help compare the minimizer $\hat{p}$ with general
elements \( p \) obeying \( \|p\| = \Delta \):
\[
0 \leq M(p) - M(\tilde{p}) \\
= \frac{1}{2} p^T H p - \frac{1}{2} \tilde{p}^T H \tilde{p} + g(p - \tilde{p}) \\
= \frac{1}{2} p^T H p - \frac{1}{2} \tilde{p}^T H \tilde{p} - \tilde{p}^T (H + \lambda I)(p - \tilde{p}) \\
= \frac{1}{2} \left( p^T (H + \lambda I)p - \lambda \|p\|^2 \right) - \frac{1}{2} \left( \tilde{p}^T (H + \lambda I)\tilde{p} - \lambda \|\tilde{p}\|^2 \right) - \tilde{p}^T (H + \lambda I)(p - \tilde{p}) \\
= \frac{1}{2} \left[ p^T (H + \lambda I)p - 2\tilde{p}^T (H + \lambda I)p + \tilde{p}^T (H + \lambda I)\tilde{p} \right] + \frac{\lambda}{2} \left[ \|\tilde{p}\|^2 - \|p\|^2 \right] \\
= \frac{1}{2}(p - \tilde{p})^T (H + \lambda I)(p - \tilde{p}) + 0.
\]
The last equation holds because \( \|p\| = \Delta = \|\tilde{p}\| \). Since the resulting inequality holds for all \( p \) with \( \|p\| = \Delta \), we deduce that \( H + \lambda I \succeq 0 \).

So to calculate the exact solution \( \tilde{p} \) for the subproblem, first try \( \lambda = 0 \): if \( H \succeq 0 \) already, and the solution \( p \) of \( Hp = -g^T \) obeys \( \|p\| \leq \Delta \), then that’s the answer. Otherwise, we define
\[
p(\lambda) = -(H + \lambda I)^{-1} g^T,
\]
noting that \( (H + \lambda I) > 0 \) for all \( \lambda > 0 \) sufficiently large, and then reduce \( \lambda \) until \( \|p(\lambda)\| = \Delta \). This is a 1D root-finding problem in the variable \( \lambda \).

**Exact Search for \( \lambda \).** This works even when \( H \) is not positive. The derivation uses the orthogonal diagonalization of \( H \):
\[
H = QQ^T, \text{ where } \Lambda = \text{diag}(\lambda_1, \lambda_2, \ldots, \lambda_n) \text{ and } \lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_n.
\]
Each column of \( Q \) is a unit vector, and they are all orthogonal: \( QQ^T = I = Q^T Q \). Hence
\[
H + \lambda I = QQ^T + \lambda QQ^T = Q(\Lambda + \lambda I)Q^T,
\]
\[
p(\lambda) = -(H + \lambda I)^{-1} g^T = -Q(\Lambda + \lambda I)^{-1} Q^T g^T = -\sum_{j=1}^{n} \frac{gq_j}{\lambda_j + \lambda} q_j.
\]
Calculating \( p(\lambda)^T p(\lambda) \) either as a double sum or as the expansion of
\[
\|p(\lambda)\|^2 = gQ(\Lambda + \lambda I)^{-1} Q^T g^T = gQ\text{diag}\left( (\lambda_j + \lambda)^{-2} \right) Q^T g^T
\]
leads to
\[
\|p(\lambda)\|^2 = \sum_{j=1}^{n} \frac{(gq_j)^2}{(\lambda_j + \lambda)^2}.
\]
The exact solution to the subproblem corresponds to the unique \( \lambda \geq -\lambda_1 \) such that \( \|p(\lambda)\|^2 = \Delta^2 \). How do we find this \( \lambda \)?

Suppose all eigenvalues are distinct, and \( gq_1 > 0 \). Then the right side converges to 0 as \( \lambda \rightarrow \infty \) and to \( +\infty \) as \( \lambda \rightarrow -\lambda_1^+ \). (Sketch.) The goal is simply to pick \( \lambda \in [\lambda_1, +\infty) \) to arrange \( \|p(\lambda)\| = \Delta \), recalling
\[
p(\lambda) = -\sum_{j=1}^{n} \frac{gq_j}{\lambda + \lambda_j} q_j.
\]
Now \( \lambda \mapsto \|p(\lambda)\| \) is very steep near \( -\lambda_1 \), so Newton’s method may work badly on this problem. A nice trick is to turn everything upside down. Newton’s method for finding a zero of \( F(\lambda) = \frac{1}{\|p(\lambda)\|} - \frac{1}{\Delta} \):

\[
\lambda^{(\ell+1)} = \lambda^{(\ell)} - \frac{F(\lambda^{(\ell)})}{F'(\lambda^{(\ell)})}.
\]

Using our earlier expression for \( \|p(\lambda)\|^2 \), we get

\[
F'(\lambda) = -\frac{1}{\|p(\lambda)\|^2} \frac{d}{d\lambda} \|p(\lambda)\| \\
= -\frac{1}{\|p(\lambda)\|^2} \frac{d}{d\lambda} \sqrt{\|p(\lambda)\|^2} \\
= -\frac{1}{\|p(\lambda)\|^2} \times \frac{1}{2\|p(\lambda)\|^2} \frac{d}{d\lambda} \|p(\lambda)\|^2 \\
= -\frac{1}{2\|p(\lambda)\|^3} \sum_{j=1}^{n} \left[ -\frac{(gq_j)^2}{(\lambda + \lambda_j)^3} \right] \\
= -\frac{1}{\|p(\lambda)\|^3} \sum_{j=1}^{n} \frac{(gq_j)^2}{(\lambda + \lambda_j)^3}
\]

Final result—Nocedal and Wright page 81:

Start with \( \lambda^{(0)} \) large enough that \( H + \lambda^{(0)} I > 0 \).

Repeat for \( \ell = 0, 1, 2, \ldots \):

Write \( \lambda = \lambda^{(\ell)} \) for simplicity.

Find \( R \) is upper triangular such that \( H + \lambda I = R^T R \). This is the Cholesky factorization (Matlab chol).

Solve for \( p \) using \( R^T R p = -g^T \).

If \( \|p\| \) sufficiently near \( \Delta \), use this vector as an approximate solution to the subproblem. Otherwise, continue:

Solve for \( z \) using \( R^T z = p \).

Use these vectors to build

\[
\lambda^{(\ell+1)} = \lambda + \left( \frac{\|p\|}{\|z\|} \right)^2 \left( \frac{\|p\| - \Delta}{\Delta} \right)
\]

Go around again.

Matlab help:

CHOL Cholesky factorization.
CHOL(X) uses only the diagonal and upper triangle of X. The lower triangular is assumed to be the (complex conjugate) transpose of the upper. If X is positive definite, then R = CHOL(X) produces an upper triangular R so that R'*R = X. If X is not positive definite, an error message is printed.

Home practice: Reconcile Newton’s method ideas above with iteration scheme above.

The Hard Case. If gradient g is orthogonal to the entire eigenspace corresponding to \( \lambda = \lambda_1 \), then the behaviour of \( \|p(\lambda)\| \) is not as described above. The correct value of \( \lambda \) to use in this case is \( \lambda = \lambda_1 \), and the corresponding \( p \) comes from adding an eigenspace component whose length is chosen appropriately to arrange \( \|p\| = \Delta \). Check the book for these details. (Or, for a rough first cut, use the Cauchy point and move rapidly to the next step, hoping the “Hard Case” is rather rare.)