MATH 612
Computational methods for equation solving and function minimization – Week # 11

F.J.S.

Spring 2014 – University of Delaware

Plan for this week

- Discuss any problems you couldn't solve from previous lectures
- We will cover Chapter 3 of the notes *Fundamentals of Optimization* by R.T. Rockafellar (University of Washington). I'll include a link in the website.
- You should spend some time reading Chapter 1 of those notes. It's full of interesting examples of optimization problems.
- Homework assignment #4 is due next Monday

UNCONSTRAINED OPTIMIZATION

Notation and problems

Data: $f : \mathbb{R}^n \to \mathbb{R}$ (**objective function**). The **feasible set** for this problem is $\mathbb{R}^n$: all points of the space are considered as possible solutions.

Global minimization problem. Find a global minimum of f:

$$x_0 \in \mathbb{R}^n \qquad f(x_0) \leq f(x) \qquad \forall x \in \mathbb{R}^n.$$

Local minimization problem. Find $x_0 \in \mathbb{R}^n$ such that there exists $\varepsilon > 0$ satisfying

$$f(x_0) \leq f(x) \qquad \forall x \in \mathbb{R}^n \quad \text{s.t.} \quad |x - x_0| < \varepsilon$$

The absolute value symbol will be used for the Euclidean norm.

Look at this formula

$$\max f(x) = -\min(-f(x))$$

Gradient and Hessian

Function $f : \mathbb{R}^n \to \mathbb{R}$. Its **gradient** vector is

$$\nabla f(x) = \left[\frac{\partial f}{\partial x_i} \right]_{i=1}^{n}.$$

In principle, we will take the gradient vector to be a column vector, so that we can dot it with a position vector x. However, in many cases points x are considered to be row vectors and then it's better to have gradients as row vectors as well.
The **Hessian** matrix of f is the matrix of second derivatives

$$(Hf)(x) = Hf(x) = \left[\frac{\partial^2 f}{\partial x_i \partial x_j} \right]_{i,j=1}^{n}.$$

When $f \in \mathcal{C}^2$, the Hessian matrix is symmetric. Notation for the Hessian is not standard.

Small o notation and more

We say that $g(x) = o(|x|^k)$ when

$$\lim_{|x|\to 0} \frac{|g(x)|}{|x|^k} = 0$$

For instance, the definition of differentiability can be written in this simple way: f is differentiable at x_0 whenever there exists a vector, which we call $\nabla f(x_0)$ such that

$$f(x) = f(x_0) + \nabla f(x_0) \cdot (x - x_0) + o(|x - x_0|).$$

When a function is of class $\mathcal{C}^2$ in a neighborhood of x_0 we can write

$$\begin{aligned} f(x) &= f(x_0) + \nabla f(x_0) \cdot (x - x_0) \\ &\quad + \tfrac{1}{2}(x - x_0) \cdot Hf(x_0)(x - x_0) + o(|x - x_0|^2) \end{aligned}$$

Descent directions

Let $x_0 \in \mathbb{R}^n$ and take $w \in \mathbb{R}^n$ as a direction for movement. Consider the function

$$0 \le t \longmapsto \varphi(t) = f(x_0 + tw).$$

Then $\varphi'(t) = \nabla f(x_0 + tw) \cdot w$, and

$$\varphi(t) = \varphi(0) + t\varphi'(0) + o(|t|) = f(x_0) + t\nabla f(x_0) \cdot w + o(|t|).$$

Then w is a descent direction when there exists an $\varepsilon > 0$ such that

$$\varphi(t) < \varphi(0) \quad t \in (0, \varepsilon) \qquad \Longleftrightarrow \qquad \nabla f(x_0) \cdot w < 0.$$

The last equivalence holds if $\nabla f(x_0) \neq 0$. The vector $w = -\nabla f(x_0)$ gives the direction of the steepest descent.

Stationary points

Let f have a local minimum at x_0. Then, for all w, $\varphi(t) = f(x_0 + tw)$ has a local minimum at $t = 0$ and

$$\varphi'(0) = \nabla f(x_0) \cdot w = 0.$$

This implies that

$$\nabla f(x_0) = 0$$

Points satisfying $\nabla f(x_0) = 0$ are called **stationary points.** Minima are stationary points, but so are maxima, and other possible points.

The sign of the Hessian at minima

Let $f \in \mathcal{C}^2(\mathbb{R}^n)$ and let x_0 be a local minimum. Then

$$\varphi(t) = \varphi(0) + \tfrac{1}{2}t^2\varphi''(0) + o(t^2) = f(x_0) + t^2\tfrac{1}{2}w \cdot Hf(x_0)w + o(t^2)$$

has a local minimum at $t = 0$ for every w. This implies that

$$w \cdot Hf(x_0)w \geq 0 \qquad \forall w \in \mathbb{R}^n,$$

that is $Hf(x_0)$ is **positive semidefinite**.

Watch out for reciprocal statements: a proof

If f is $\mathcal{C}^2$, $\nabla f(x_0) = 0$ and $Hf(x_0)$ is positive definite (not semidefinite!), then f has a local minimum at x_0.
Proof. For $x \neq x_0$,

$$f(x) = f(x_0) + \underbrace{\tfrac{1}{2}(x - x_0) \cdot Hf(x_0)(x - x_0)}_{=g(x)>0} + \underbrace{h(x)}_{=o(|x-x_0|)^2}$$

On the other hand,

$$w \cdot Hf(x_0)w \geq c|w|^2 \qquad \forall w \in \mathbb{R}^n, \qquad \text{with } c > 0$$

(why?) and therefore we can find $\varepsilon > 0$ such that

$$|h(x)| \leq \tfrac{c}{4}|x - x_0|^2 < |g(x)| \qquad 0 < |x - x_0| < \varepsilon,$$

which proves that x_0 is a **strict local minimum.**

Watch out for reciprocal statements: counterexamples

If $\nabla f(x_0) = 0$ and $Hf(x_0)$ is positive semidefinite, things can go in several different ways. In one variable

$$\psi(t) = t^3$$

has $\psi'(0) = 0$ (stationary point), $\psi''(0) = 0$ (positive semidefinite), but there's no local minimum at $t = 0$.
In two variables

$$f(x, y) = x^2 + y^3$$

has $\nabla f(0, 0) = 0$,

$$Hf(0, 0) = \begin{bmatrix} 2 & 0 \\ 0 & 0 \end{bmatrix} \qquad \text{positive semidefinite}$$

and no local minimum at the origin.

SIMPLE FUNCTIONALS

Linear functionals

Doing unconstrained minimization for linear functionals

$$f(x) = x \cdot b + c$$

is not really an interesting problem. This is why:

$$\nabla f(x) = b, \qquad Hf(x) = 0.$$

Only constant functionals have minima, but all points are minima in that case. *Note, however, that we will deal with linear functionals for constrained optimization problems.*

Quadratic functionals

Let A be a symmetric matrix, $b \in \mathbb{R}^n$ and $c \in \mathbb{R}$. We then define

$$f(x) = \frac{1}{2} x \cdot Ax - x \cdot b + c$$

and compute

$$\nabla f(x) = Ax - b, \qquad Hf = A.$$

- Stationary points are solutions to $Ax = b$.
- Local minima exist only when A is positive semidefinite.
- If A is positive definite, then there is only one stationary point, which is a global minimum. (Proof in the next slide.)

Quadratic functionals (2)

If $Ax_0 = b$ and A is positive definite, then

$$f(x_0) = f(x_0) + \frac{1}{2}(x - x_0) \cdot A(x - x_0) > f(x_0), \qquad x \neq x_0,$$

because there's no remainder in Taylor's formula of order two. What happens when A is positive semidefinite? On of these two possibilities:

- There are no critical points ($Ax = b$ is not solvable). We can (how?) then find x_* such that $Ax_* = 0$ and $x_* \cdot b > 0$. Using vectors tx_* for $t \to \infty$, we can see that f is unbounded below
- There is a subspace of global minima (all critical points = all solutions to $Ax = b$).

A control-style quadratic minimization problem

For a positive semidefinite matrix W, an invertible matrix C, and suitable matrices and vectors D, b and b, we minimize the functional:

$$f(u) = \frac{1}{2} x \cdot Wx - x \cdot b + |u|^2, \qquad \text{where} \qquad Cx = Du + d$$

As an exercise, write this functional as a functional in the variable u alone (in the jargon of control theory, x is a state variable) and find the gradient and Hessian of f.

CONVEXITY

Convex functions (functionals)

A function $f : \mathbb{R}^n \to \mathbb{R}$ is **convex** when

$$f(\,(1-\tau)x_0 + \tau\, x_1) \le (1-\tau)f(x_0) + \tau\, f(x_1) \qquad \forall \tau \in (0,1), \qquad \forall x_0, x_1 \in \mathbb{R}^n.$$

It is **scrictly convex** when

$$f(\,(1-\tau)x_0 + \tau\, x_1) < (1-\tau)f(x_0) + \tau\, f(x_1) \qquad \forall \tau \in (0,1), \qquad \forall x_0 \ne x_1 \in \mathbb{R}^n.$$

A function f is **concave** when $-f$ is convex.

Confusing? Easy to remember

- In undergraduate textbooks, convex is said concave up, and concave is said concave down.
- Grown-ups (mathematicians, scientists, engineers) always use convex with this precise meaning. There's no ambiguity. Everybody uses the same convention.
- x^2 is convex. Repeat yourself this many times.

Line/segment convexity

Take $x_0 \neq x_1$ and the segment

$$[0,1] \ni \tau \longmapsto x(\tau) = (1-\tau)x_0 + \tau x_1 .$$

If the function f is convex, then the one dimensional function

$$\varphi(t) = f(x(t))$$

is also convex:

$$\varphi(t) = \varphi((1-t)0+t1) \leq (1-t)\varphi(0)+t\varphi(1) = (1-t)f(x_0)+tf(x_1).$$

This segment-convexity is equivalent to the general concept of convexity. In other words, a function is convex if and only if it is convex by segments for all segments.

Jensen's inequality

A function f is convex if and only if for all $k \geq 1$, $x_0, \ldots, x_k \in \mathbb{R}^n$, and $\tau_0 + \ldots + \tau_k = 1$, $\tau_j \geq 0$,

$$f\left(\tau_0 x_0 + \tau_1 x_1 + \ldots + \tau_k x_k\right) \leq \tau_0 f(x_0) + \tau_1 f(x_1) + \ldots + \tau_k f(x_k)$$

The expression

$$\sum_{j=0}^{k} \tau_j x_j \qquad \text{where} \qquad \tau_j \geq 0, \forall j \qquad \sum_{j=0}^{k} \tau_j = 1$$

is called a **convex combination** of the points $x_0, \ldots, x_k$. The set of all convex combinations of the points $x_0, \ldots, x_k$ is called the **convex hull** of the points $x_0, \ldots, x_k$.

Jensen's inequality: proof by induction

The case $k = 1$ is just the definition with $\tau_0 = 1 - \tau$ and $\tau_1 = \tau$.
For a given k

$$
\begin{aligned}
f(\sum_{j=0}^{k} \tau_j x_j) &= f\Big(\tau_0 x_0 + (1-\tau_0)(\sum_{j=1}^{k} \frac{\tau_j}{1-\tau_0} x_j)\Big) \\
&\le \tau_0 f(x_0) + (1-\tau_0) f\Big(\sum_{j=1}^{k} \frac{\tau_j}{1-\tau_0} x_j\Big) \\
&\le \tau_0 f(x_0) + (1-\tau_0) \sum_{j=1}^{k} \frac{\tau_j}{1-\tau_0} f(x_j) \qquad \text{Note: } \Big(\textstyle\sum_{j=1}^{k} \frac{\tau_j}{1-\tau_0} = 1\Big) \\
&= \sum_{j=0}^{k} \tau_j f(x_j).
\end{aligned}
$$

(Note that if $\tau_0 = 1$ there's nothing to prove.)

An argument

Assume that f is convex. If there exist x_0 and $\varepsilon > 0$ such that

$$f(x_0) \leq f(x_0 + \varepsilon w) \qquad \forall w \in \mathbb{R}^n \text{ with } |w| = 1,$$

then

$$\begin{aligned} f(x_0) \leq f(x_0 + \varepsilon w) &= f\Big(\tfrac{\varepsilon}{t+\varepsilon} x_0 + \tfrac{t}{t+\varepsilon}(x_0 + (t+\varepsilon)w)\Big) \\ &\leq \tfrac{\varepsilon}{t+\varepsilon} f(x_0) + \tfrac{t}{t+\varepsilon} f(x_0 + (t+\varepsilon)w)\Big), \end{aligned}$$

and

$$\tfrac{t}{t+\varepsilon} f(x_0) = \Big(1 - \tfrac{\varepsilon}{t+\varepsilon}\Big) f(x_0) \leq \tfrac{t}{t+\varepsilon} f(x_0 + (t+\varepsilon)w)$$

which implies

$$f(x_0) \leq f(x_0 + tw) \qquad \forall t \geq \varepsilon \qquad \forall w \text{ with } |w| = 1.$$

A conclusion

The previous argument (and some minor additional work) shows that for a convex function, **any local minimum is a global minimum.**

This does not mean that convex functions have global minima. For instance

$$e^{-x_1} + e^{-x_2} + \ldots + e^{-x_n}$$

is strictly convex (why?) and does not have minimum value.

Another result

If $x_0, \ldots, x_k$ are minima of a convex function

$$f(x_0) = \ldots = f(x_k) \leq f(x) \qquad \forall x \in \mathbb{R}^n,$$

then with for any convex combination

$$c \leq f(\sum_{j=0}^{k} \tau_j x_j) \leq \sum_{j=0}^{k} \tau_j f(x_j) = c \sum_{j=0}^{k} \tau_j = c.$$

Therefore the convex hull of a set of minima contains also minima.

Scrict convexity brings uniqueness

If $x_0 \neq x_1$ are two global minima of a scritcly convex function

$$c = f(x_0) = f(x_1) \leq f(x) \qquad \forall x \in \mathbb{R}^n,$$

then

$$f(\tfrac{1}{2}x_0 + \tfrac{1}{2}x_1) < \tfrac{1}{2}f(x_0) + \tfrac{1}{2}f(x_1) = c,$$

which contradicts our hypothesis on having found two minima. The strict inequality does not happen when $x_0 = x_1$ and this shows uniqueness.

CONVEXITY OF SMOOTH FUNCTIONS

Convexity and tangent line

Let $\varphi : \mathbb{R} \to \mathbb{R}$. Then φ is convex if and only if

$$\varphi(t) \geq \underbrace{\varphi(\tau) + \varphi'(\tau)(t-\tau)}_{\text{tangent line at } \tau} \qquad \forall t, \tau. \tag{1}$$

Proof. Take $t > \tau$. Then

$$\begin{aligned} \varphi \text{ is convex} \quad &\Longleftrightarrow \quad \varphi' \text{ is non-decreasing} && \text{(HW4)} \\ &\Longrightarrow \quad \frac{\varphi(t)-\varphi(\tau)}{t-\tau} \geq \varphi'(\tau) && \text{(MVT)} \end{aligned}$$

(A similar argument works for $\tau > t$.) Using (1) for the pairs (t, τ) and (τ, t) proves that

$$(\varphi'(\tau) - \varphi'(t))(t-\tau) \leq 0 \qquad \forall t, \tau,$$

that is, φ' is non-decreasing.

Convexity and tangent plane

Let $f : \mathbb{R}^n \to \mathbb{R}$ be differentiable at every point. Then f is convex if and only if

$$f(y) \geq \underbrace{f(x) + \nabla f(x) \cdot (y - x)}_{\text{tangent plane at } x} \qquad \forall x, y \in \mathbb{R}^n. \tag{2}$$

Proof. Let $\mathbb{R} \ni t \mapsto z(t) = x + t(y - x)$, and $\varphi(t) = f(z(t))$. If f is convex, then φ is convex and by the one-dimensional result

$$\varphi(1) \geq \varphi(0) + \varphi'(0) \qquad \text{that is, (\textbf{??}).}$$

If (**??**) holds, then

$$\varphi(t) = f(z(t)) \geq f(z(\tau)) + \nabla f(z(\tau)) \cdot \underbrace{(z(t) - z(\tau))}_{(t-\tau)(y-x)} = \varphi(\tau) + \varphi'(\tau)(t - \tau)$$

and φ is convex. Finally, line-convexity implies convexity.

Corollary: stationary points of convex functions

If f is convex and differentiable and x_0 is a stationary point

$$\nabla f(x_0) = 0,$$

then x_0 is a global minimum.

Proof. We know that

$$f(x) \geq f(x_0) + \nabla f(x_0) \cdot (x - x_0) = f(x_0) \qquad \forall x \in \mathbb{R}^n,$$

so this is the proof.

If f is strictly convex and differentiable, then there is at most one stationary point which will be the only global minimum.

Strict convexity

Let $f : \mathbb{R}^n \to \mathbb{R}$ be differentiable at every point. Then f is convex if and only if

$$f(y) > \underbrace{f(x) + \nabla f(x) \cdot (y - x)}_{\text{tangent plane at } x} \qquad \forall x, y \in \mathbb{R}^n, \qquad y \neq x$$

The argument is very similar. Use first that for functions of one-variable, strict convexity of φ is equivalent to φ' being increasing. Then use line parametrizations to go from n dimensions to one dimension.

Convexity and second derivative

Let $\varphi : \mathbb{R} \to \mathbb{R}$ be twice differentiable.

- φ is convex if and only if $\varphi''(t) \geq 0$ for all t. (Proof. φ is convex iff φ' is non-decreasing!)
- If $\varphi''(t) > 0$ for all t, then φ is convex. (Proof. φ' is increasing!)
- The function $\varphi(t) = t^4$ is strictly convex, but $\varphi'(0) = 0$.

Example. $\exp(t)$ is strictly convex.

Let $f : \mathbb{R}^n \to \mathbb{R}$ be twice differentiable. Then f is convex if and only if

$$Hf(x) \quad \text{is positive semidefinite} \qquad \forall x.$$

If $Hf(x)$ is positive definite for all x, then f is strictly convex.

Proof. Take $z(t) = x + t(y - x)$ and $\varphi(t) = f(z(t))$. Then f is convex if and only if all the functions φ are convex (for arbitrary choice of x and y), if and only if

$$\varphi''(\tau) = (y - x) \cdot Hf(x)(y - x) \geq 0 \qquad \forall x, y.$$

The strictly convex case is similar.

Example and counter-example. The function $\exp(b \cdot x)$, with $b_i \neq 0$ for all i is strictly convex. The function $x_1^4 + \ldots + x_n^4$ is strictly convex but has vanishing Hessian at the origin.

DESCENT METHODS

A problem and two ideas

Problem. Find the unconstrained minimum of a (convex) function f.

Goal of the method. Produce a sequence of points reducing the value of f:

$$f(x^\nu) > f(x^{\nu+1}) \qquad \forall \nu.$$

Find a descent direction. For each ν, find a descent direction w^ν, that is,

$$f(x^\nu + tw^\nu) < f(x^\nu) \qquad \text{for } 0 > t > \varepsilon.$$

If f is differentiable: $\nabla f(x^\nu) \cdot w^\nu < 0$. The **steepest descent** method consists of taking $w^\nu = -\nabla f(x^\nu)$.

Do a line search. Find a value $t^\nu > 0$ ensuring that

$$f(x^\nu + t^\nu w^\nu) < f(x^\nu)$$

and define $x^{\nu+1} = x^\nu + t^\nu w^\nu$.

Exact line search

We have the point x^ν and the descent direction w^ν. We then define the function

$$[0,\infty) \ni t \longmapsto \varphi(t) = \varphi^\nu(t) = f(x^\nu + t\,w^\nu).$$

This function decreases near 0. If f is convex, this function is convex and: (a) either has a minimum at some $t > 0$, (b) or is unbounded below (so is the original function); (c) or decreases to a limit as $t \to \infty$.
We assume that we are in the (a) case. We then solve the one-dimensional minimization problem:

$$\text{find } t^\nu > 0 \text{ such that} \quad \varphi(t^\nu) \le \varphi(t) \qquad \forall t \in [0,\infty).$$

Exact line search (2)

If f is convex, so is $\varphi(t) = f(x^\nu + tw^\nu)$. If f is differentiable, we only need to look for a stationary point

$$\varphi'(t) = 0 \qquad \Longleftrightarrow \qquad \nabla f(x^\nu + tw^\nu) \cdot w^\nu = 0.$$

This is a non-linear equation of a single variable. It can be solved with Newton iterations:

$$\tau_{k+1} = \tau_k - \frac{\varphi(\tau_k)}{\varphi'(\tau_k)},$$

at the cost of one evaluation of f and one of ∇f at each iteration.

Backtracking

If f is differentiable (actually convexity is enough but we won't say why), then

$$\frac{1}{\tau}(\varphi(\tau) - \varphi(0)) \stackrel{\tau \to 0}{\longrightarrow} \varphi'(0) < \beta\varphi'(0) < 0,$$

for $0 < \beta < 1$ (chosen parameter). We then look for $0 < \tau < 1$ satisfying

$$\varphi(\tau) - \varphi(0) < \tau\,\beta\,\varphi'(0) < 0.$$

The value τ is found by considering $\tau = \gamma^k$ as k grows, where $0 < \beta < \gamma < 1$ is another desing parameter.

Backtracking (2)

for $\nu \geq 0$
    find a descent direction w
    $\phi_0 = f(x)$
    $\psi_0 = \nabla f(x) \cdot w$
    $\tau = \gamma$
    $\varphi_1 = f(x + \tau\, w)$
    while $\varphi_1 \geq \varphi_0 + \tau\beta\psi_0$
        $\tau = \tau\gamma$
        $\varphi_1 = f(x + \tau\, w)$
    end
    $x = x + \tau w$
    stopping criterion
end

Steepest descent for quadratic functions

The objective function is

$$f(x) = \frac{1}{2} x \cdot Ax - x \cdot b,$$

where A is symmetric positive definite. We know that

$$\nabla f(x) = Ax - b.$$

At the iteration ν, we have x^ν and compute the descent direction

$$w^\nu = -\nabla f(x^\nu) = b - Ax^\nu = r^\nu.$$

Therefore, the descent direction is the residual.

Follow me!

$$
\begin{aligned}
\varphi(t) &= f(x^\nu + tw^\nu) \\
&= \tfrac{1}{2}(x^\nu + tw^\nu)\cdot A(x^\nu + tw^\nu) - (x^\nu + tw^\nu)\cdot b \\
&= f(x^\nu) + tw^\nu \cdot (Ax^\nu - b) + \tfrac{1}{2}t^2 w^\nu \cdot Aw^\nu \\
&= f(x^\nu) - t|w^\nu|^2 + \tfrac{1}{2}t^2 w^\nu \cdot Aw^\nu .
\end{aligned}
$$

The minimum for this quadratic functional is attained at

$$
t = \frac{|w^\nu|^2}{w^\nu \cdot Aw^\nu} = \frac{|r^\nu|^2}{r^\nu \cdot Ar^\nu}.
$$

We recover the Steepest Descent method for the positive definite system $Ax = b$.

NEWTON

First approach: stationary points

For a convex function, any stationary point

$$\nabla f(x) = 0$$

is a global minimum. We then find roots of

$$F(x) = 0, \qquad F = \nabla f : \mathbb{R}^n \to \mathbb{R}^n.$$

Newton's iteration for systems is defined as

$$x^{\nu+1} = x^{\nu} - \nabla F(x^{\nu})^{-1} F(x^{\nu}),$$

where

$$\nabla F(x)_{ij} = \frac{\partial F_i}{\partial x_j}$$

In our case

$$F = \nabla f, \qquad \nabla F = Hf.$$

First approach: stationary points

The implementation form is

for $\nu \geq 1$

$b = F(x)$

$A = \nabla F(x)$

Solve $Aw = b$

$x = x - w$

Stopping criterion

end

For stationary points, susbtitute $F = \nabla f$, $\nabla F = Hf$.

Second approach: quadratic approximation

Given x^ν consider the quadratic Taylor approximation

$$q(x) = f(x^\nu) + \nabla f(x^\nu) \cdot (x - x^\nu) + \tfrac{1}{2}(x - x^\nu) \cdot Hf(x^\nu)(x - x^\nu).$$

It attains its minimum at

$$x = x^\nu - Hf(x^\nu)^{-1} \nabla f(x^\nu).$$

We then move to the minimum for the quadratic approximation and repeat the process. What we get is **exactly** Newton's method to find stationary points.

Newton descent method

In both algorithms above, we found $x^{\nu+1} = x^{\nu} + w^{\nu}$, where

$$w^{\nu} = -Hf(x^{\nu})^{-1}\nabla f(x^{\nu}).$$

Note that

$$w^{\nu} \cdot \nabla f(x^{\nu}) = -\nabla f(x^{\nu}) \cdot Hf(x^{\nu})^{-1}\nabla f(x^{\nu}) < 0,$$

so w^{ν} is a descent direction. Newton method for optimization consists of combining the Newton choice of descent direction with some kind of line search. Then the iteration is

$$x^{\nu+1} = x^{\nu} + t^{\nu} w^{\nu} = x^{\nu} - t^{\nu} Hf(x^{\nu})^{-1}\nabla f(x^{\nu}).$$

Newton descent with backtracking line search

for $\nu \geq 1$

$b = \nabla f(x)$

$A = Hf(x)$

$w = A^{-1}b$

$\varphi_0 = f(x)$, $\psi_0 = w \cdot b$

$\tau = \gamma$

$\varphi_1 = f(x + \tau w)$

while $\varphi_1 > \varphi_0 + \tau\beta\psi_0$

$\tau = \tau\gamma$

$\varphi_1 = f(x + \tau w)$

end

$x = x + \tau w$

stopping criterion

end

CONVERGENCE

Strictly convex functions

Let $f : \mathbb{R}^n \to \mathbb{R}$ be:

- strictly convex
- with bounded level sets

$$\{x \in \mathbb{R}^n \ : \ f(x) \leq \alpha\} \text{ bounded, for all } \alpha$$

We use a descent method with **exact line search** and:

- steepest descent (assuming $f \in \mathcal{C}^1$)
- Newton search (assuming $f \in \mathcal{C}^2$)
- any choice of descent that is a continuous function of the point

Then the descent method converges to the only global minimum of f.

Modifications of the theorem

- If we relax strict convexity of f, Newton's method is not applicable, since it's based on

$$\nabla f(x^\nu) \cdot w^\nu = -\nabla f(x^\nu) \cdot Hf(x^\nu) \nabla f(x^\nu) < 0$$

  which means (note how the Hessian has to be invertible) that we need $Hf(x)$ to be positive definite.
- If we relax convexity, there's still some kind of convergence. With steepest descent or any other continuous choice of descent direction, the sequence x^ν might not converge, but it is bounded and all its accumulation points are critical points.