9.1 *Minimizing a quadratic function.* Consider the problem of minimizing a quadratic function:

$$\text{minimize} \quad f(x) = (1/2)x^T P x + q^T x + r,$$

where $P \in \mathbf{S}^n$ (but we do not assume $P \succeq 0$).

(a) Show that if $P \not\succeq 0$, *i.e.*, the objective function f is not convex, then the problem is unbounded below.

(b) Now suppose that $P \succeq 0$ (so the objective function is convex), but the optimality condition $Px^\star = -q$ does not have a solution. Show that the problem is unbounded below.

(a). $P \not\succeq 0. \Rightarrow \exists\, v$ s.t. $v^T P v < 0.$

let $x = tv$. $t \in \mathbb{R}$.

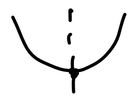
then $f(x) = \frac{1}{2} t^2 (v^T P v) + t\, q^T v + r$

$$= \underbrace{(v^T P v/2)}_{<0} t^2 + (q^T v) t + r$$

when $t \to -\infty$. $/ +\infty$. $f(x) \to -\infty$. $\therefore$ unbounded below.

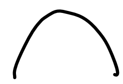
scalar analogy:

$ax^2 + bx + c$



$a > 0$

$x^* = -\frac{b}{2a}$



$a < 0$

unbounded below

(b) $Px^* = -q$ has no solution. q is not in the range of P.

write q as $\bar{q} + v$. where

$\underbrace{\bar{q} = Pu}_{\text{projection}}$ and $\boxed{\underbrace{v^T (Pw) = 0.}_{\text{orthogonal to range}(P)}\ \forall w}$

if we let $x = tv$. $t \in \mathbb{R}$.

$f(x) = q^T t v + r = t q^T v + r$

$= \underline{t \bar{q}^T} v + t v^T v + r$

$= t \cdot v^T v + r \to -\infty$ when $t \to -\infty$. (since $v \neq 0$)

Scalar analogy

$a \geq 0$. but $a = 0$.

($x^* = -b/2a$ is not a solution)

9.6 *Quadratic problem in* $\mathbf{R}^2$. Verify the expressions for the iterates $x^{(k)}$ in the first example of §9.3.2.

Our first example is very simple. We consider the quadratic objective function on $\mathbf{R}^2$

$$f(x) = \frac{1}{2}(x_1^2 + \gamma x_2^2),$$

We apply the gradient descent method with exact line search, starting at the point $x^{(0)} = (\gamma, 1)$. In this case we can derive the following closed-form expressions for the iterates $x^{(k)}$ and their function values (exercise 9.6):

$$x_1^{(k)} = \gamma\left(\frac{\gamma-1}{\gamma+1}\right)^k, \qquad x_2^{(k)} = \left(-\frac{\gamma-1}{\gamma+1}\right)^k, \qquad (*)$$

Gradient descent with exact line search.

until converge:

$$t^{(k)} = \underset{t \geq 0}{\operatorname{argmin}}\ f(x^{(k)} - t\nabla f(x^{(k)})),$$

$$x^{(k+1)} = x^{(k)} - t^{(k)} \nabla f(x^{(k)})$$

We use M.I. to prove $(*)$:

1°. Base case: $x_1^{(0)} = \gamma \cdot (\quad)^0 = \gamma$

$x_2^{(0)} = (\quad)^0 = 1.$ ✓

2°. Assume we know that $x_1^{(k)} = \gamma\left(\frac{\gamma-1}{\gamma+1}\right)^k$, $x_2^{(k)} = \left(-\frac{\gamma-1}{\gamma+1}\right)^k$.

We find $x^{(k+1)}$.

$$\nabla f(x) = \begin{bmatrix} x_1 \\ \gamma x_2 \end{bmatrix}$$

$$\therefore\ x^{(k)} - t\nabla f(x^{(k)}) = \begin{bmatrix} (1-t)\, x_1^{(k)} \\ (1-t\gamma)\, x_2^{(k)} \end{bmatrix} = \left(\frac{\gamma-1}{\gamma+1}\right)^k \begin{bmatrix} (1-t)\gamma \\ (1-t\gamma)(-1)^k \end{bmatrix}$$

$\therefore$ to minimize $f(x^{(k)} - t\nabla f(x^{(k)}))$,

is to minimize: $(1-t)^2\gamma^2 + \gamma(1-t\gamma)^2((-1)^k)^2$

$$= (1-t)^2\gamma^2 + \gamma(1-t\gamma)^2.$$

$$-2\gamma^2(1-t) - 2\gamma^2(1-\gamma t) = 0$$

$$1-t+1-\gamma t = 0 \Rightarrow t^{(k)} = \frac{2}{1+\gamma}$$

$$\therefore\ x^{(k+1)} = x^{(k)} - t^{(k)}\nabla f(x^{(k)}) = \begin{bmatrix} (1-t^{(k)})\, x_1^{(k)} \\ (1-t^{(k)}\cdot\gamma)\, x_2^{(k)} \end{bmatrix}$$

$$= \begin{bmatrix} \left(\frac{\gamma-1}{1+\gamma}\right) x_1^{(k)} \\ \left(\frac{1-\gamma}{1+\gamma}\right) x_2^{(k)} \end{bmatrix} = \begin{bmatrix} \gamma\cdot\left(\frac{\gamma-1}{\gamma+1}\right)^{k+1} \\ \left(\frac{1-\gamma}{1+\gamma}\right)^{k+1} \end{bmatrix}. \quad ✓$$

5. Consider the function

$$f(x) = \|x\|^{2+\beta}, \quad \text{where} \quad \beta \geq 0$$

We use the gradient method with constant stepsize s to minimize f. For which values of $x^{(0)}$ and s does the method converge to $x^* = 0$?

First convince yourself that if $\|x^{(1)}\| < \|x^{(0)}\|$ then $\|x^{(k+1)}\| < \|x^{(k)}\|$ and if $\|x^{(1)}\| \geq \|x^{(0)}\|$ then $\|x^{(k+1)}\| \geq \|x^{(k)}\|$.

Is ∇f Lipschitz, i.e., does there exist a finite L such that

$$\|\nabla f(x) - \nabla f(y)\| \leq L\|x - y\|?$$

Recall: convergence analysis for fixed-step gradient descent:

f is convex. differentiable. ∇f is Lipschitz continuous.

Then gradient descent with fixed step size $t \leq \frac{1}{L}$ will have:

$$f(x^{(k)}) - f(x^*) \leq \frac{\|x^{(0)} - x^*\|_2}{2tk}$$

[Convergence rate is $O(1/k)$. to get $f(x^{(k)}) - f(x^*) \leq \varepsilon$, we need $k \in O(1/\varepsilon)$ iterations.]

· Is ∇f Lipschitz continuous?

$$f(x) = (x_1^2 + x_2^2 + \cdots + x_n^2)^{1+\frac{\beta}{2}}$$

$$\nabla f = \begin{bmatrix} (1+\beta/2)(x_1^2+x_2^2+\cdots+x_n^2)^{\beta/2} \cdot 2x_1 \\ (1+\beta/2)(x_1^2+x_2^2+\cdots+x_n^2)^{\beta/2} \cdot 2x_2 \\ \vdots \\ (1+\beta/2)(x_1^2+x_2^2+\cdots+x_n^2)^{\beta/2} \cdot 2x_n \end{bmatrix} = (2+\beta)\|x\|^{\beta} \cdot x$$

Does there exist $L > 0$. s.t. $\|\nabla f(x) - \nabla f(y)\| \leq L\|x-y\|$, $\forall x, y$.

$$\Leftrightarrow \quad \|(2+\beta)\|x\|^{\beta} x - (2+\beta)\|y\|^{\beta} y\| \leq L\|x-y\|$$

$$\Leftrightarrow \quad (2+\beta)\,\|\,\|x\|^{\beta} x - \|y\|^{\beta} y\| \leq L\|x-y\|$$

let $x = -y$. $\quad (2+\beta)\,\|\,\|x\|^{\beta}(2x)\| \leq L\|(2x)\|$

$$\Leftrightarrow \quad (2+\beta)\|x\|^{\beta} \leq L \quad \text{for all } x.$$

Apparently, there's no such finite L. ∇f is not Lipschitz cont.

$$x^{(k+1)} = x^{(k)} - s\nabla f(x^{(k)}) = x^{(k)} - s(2+\beta)\|x^{(k)}\|^{\beta} x^{(k)}$$

$$= x^{(k)} \cdot \left(1 - s(2+\beta)\|x^{(k)}\|^{\beta}\right)$$

We want to show that if $\|x^{(1)}\| < \|x^{(0)}\|$. then $\|x^{(k+1)}\| < \|x^{(k)}\|$ for all k. We show this by M.I.

Assume we have $\|x^{(k)}\| < \|x^{(k-1)}\|$ for $k \leq n$. then for n+1.

we want : $\|x^{(n+1)}\| / \|x^{(n)}\| = |1 - s(2+\beta)\|x^{(n)}\|^{\beta}| \; (< 1)$

$\iff \quad 0 < s(2+\beta)\|x^{(n)}\|^{\beta} < 2$

By hypothesis, $\|x^{(n)}\| < \|x^{(n-1)}\|$

$\iff \quad 0 < s(2+\beta)\|x^{(n-1)}\|^{\beta} < 2$.

$\therefore \; s(2+\beta)\|x^{(n)}\|^{\beta} < s(2+\beta)\|x^{(n-1)}\|^{\beta} < 2$

Also. let $s>0$. $s(2+\beta)\|x^{(n)}\|^{\beta} > 0$.

$\therefore \; \|x^{(n+1)}\| < \|x^{(n)}\|$ ✓

Thus. for $x^{(k)}$ to converge to 0. we must have $\|x^{(1)}\| < \|x^{(0)}\|$.

$\iff \quad s(2+\beta)\|x^{(0)}\|^{\beta} < 2 \qquad (*)$.

Now we argue this is sufficient. i.e. for β, s, and $x^{(0)}$ satisfying $(*)$. $\Rightarrow x^{(k)} \to 0$.

Since $\{\|x^{(k)}\|\}$ is non-increasing. and ≥ 0. it has a limit, c.

if $c>0$. then $\lim_{k\to+\infty} \dfrac{\|x^{(k+1)}\|}{\|x^{(k)}\|} = 1$

However, we have $\lim_{k\to+\infty} \dfrac{\|x^{(k+1)}\|}{\|x^{(k)}\|} = \lim_{k\to+\infty} |1 - s(2+\beta)\|x^{(k)}\|^{\beta}|$

$= |1 - s(2+\beta)c^{\beta}| = 1 \Rightarrow s(2+\beta)c^{\beta} = 2$

On the other hand $c < \|x^{(0)}\|$. and from $(*)$:

$s(2+\beta)c^{\beta} < s(2+\beta)\|x^{(0)}\|^{\beta} < 2$

$\therefore$ contradiction.

$\therefore$ c must be zero. $\therefore$ Converge as long as $(*)$ is satisfied