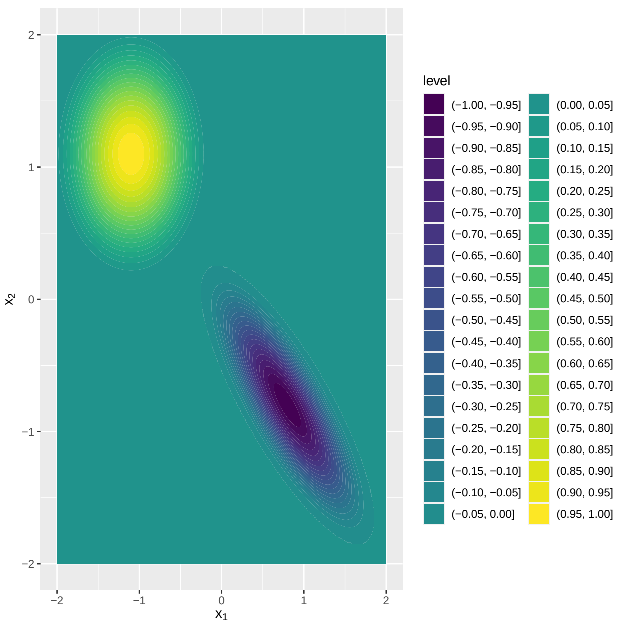
Multivariate Optimization 2

Solution 1: Gradient Descent

(a)
```
library(ggplot2)

c1 = c(-1.1,1.1)
c2 = c(0.8, -0.8)

S2 = matrix(c(1.1, -0.9, -0.9, 1.1), nrow = 2)
S2_inv = solve(S2)

rho <- function(u) {ifelse(abs(u) < 1, (1 - u^2)^2, 0)}

dist1 <- function(x) {sqrt((x - c1) %*% (x -c1))}
dist2 <- function(x) {sqrt((x - c2)  %*% S2_inv %*% (x - c2))}

f <- function(x) {rho(dist1(x)) - rho(dist2(x))}

x = seq(-2, 2, by=0.01)
xx = expand.grid(X1 = x, X2 = x)

fxx = apply(xx, 1, f)
df = data.frame(xx = xx, fxx = fxx)

cont_plot = ggplot() +
    geom_contour_filled(data = df, aes(x = xx.X1, y = xx.X2, z = fxx),
                        binwidth = 0.05) +
  xlab(expression(x[1])) +
  ylab(expression(x[2]))
cont_plot
```



(b) First we analyze $\rho(u)$ for $|u| < 1 : (1-u^2)^2 = 0 \iff (1-u^2) = 0 \iff u^2 = 1 \Rightarrow \rho(u) \neq 0$ for $u^2 < 1$ and $\rho(u) = 0$ for $u^2 \geq 1$.
We can check this condition for both squared distances around the centers $\mathbf{c}_1, \mathbf{c}_2$:

(i) $\|\mathbf{x}-\mathbf{c}_1\|_{S_1}^2 < 1 \iff \|\mathbf{x}-\mathbf{c}_1\|_2^2 < 1$ (unit circle around $\mathbf{c}_1$)

(ii) $\|\mathbf{x}-\mathbf{c}_2\|_{S_2}^2 = (\mathbf{x}-\mathbf{c}_2)^\top \begin{pmatrix} 1.1 & -0.9 \\ -0.9 & 1.1 \end{pmatrix}^{-1} (\mathbf{x}-\mathbf{c}_2) < 1$ (ellipse around $\mathbf{c}_2$)
In order to find the smallest enclosing circle of the ellipse we can use the eigendecomposition of S_2 :
$\det(S_2 - \lambda\mathbf{I}) = 0 \iff \det\begin{pmatrix} 1.1-\lambda & -0.9 \\ -0.9 & 1.1-\lambda \end{pmatrix} = 0 \iff \lambda^2 - 2.2\lambda + 0.4 = 0 \iff \lambda_1 = 2.0, \lambda_2 = 0.2$
$\Rightarrow$ Eigenvalues μ_1, μ_2 of S_2^{-1} are $\mu_i = 1/\lambda_i$.
With this we get
$\|\mathbf{x}-\mathbf{c}_2\|_{S_2}^2 < 1 \iff (\mathbf{x}-\mathbf{c}_2)^\top \mathbf{V}^\top \begin{pmatrix} 5 & 0 \\ 0 & 0.5 \end{pmatrix} \mathbf{V}(\mathbf{x}-\mathbf{c}_2) < 1$ with $|\det \mathbf{V}| = 1$.
$\Rightarrow$ the circle around $\mathbf{c}_2$ with radius $\sqrt{1/0.5} = \sqrt{2}$ encloses the ellipse.
$\|\mathbf{c}_2 - \mathbf{c}_1\|_2 = \sqrt{2 \cdot 1.9^2} \approx 2.69 > 1 + \sqrt{2} \approx 2.41 \Rightarrow$ the circles can not intersect
$\Rightarrow$ the unit circle around $\mathbf{c}_1$ and the ellipse around $\mathbf{c}_2$ can not intersect $\Rightarrow$ only $\rho(\|\mathbf{x}-\mathbf{c}_1\|_{S_1})$ or $\rho(\|\mathbf{x}-\mathbf{c}_2\|_{S_2})$ can be non-zero for a given $\mathbf{x} \in \mathbb{R}^2$.

(c) Because of b) we know that we can treat $\rho(\|\mathbf{x}-\mathbf{c}_1\|_{S_1})$ and $\rho(\|\mathbf{x}-\mathbf{c}_2\|_{S_2})$ independently. Also it follows from $\rho(u) \geq 0\ \forall u \in \mathbb{R}, w_1 > 0$ and $w_2 < 0$ that the global minimum must be in $\{\mathbf{x} \in \mathbb{R}^2 |\ \|\mathbf{x}-\mathbf{c}_2\|_{S_2}^2 < 1\}$
$\frac{\partial}{\partial \mathbf{x}}\rho(\|\mathbf{x}-\mathbf{c}_2\|_{S_2}) = 2(1-\|\mathbf{x}-\mathbf{c}_2\|_{S_2}^2)\cdot(-2)\cdot(\mathbf{x}-\mathbf{c}_2)^\top S_2^{-1} \stackrel{!}{=} \mathbf{0} \Rightarrow$ either $\|\mathbf{x}-\mathbf{c}_2\|_{S_2}^2 = 1$ (which is the boundary) or $\mathbf{x} = \mathbf{c}_2$.
Since $-\rho(1) = 0$ and $-\rho(\|\mathbf{c}_2 - \mathbf{c}_2\|) = -1 < 0$ it follows that the global minimum must be $\mathbf{x} = \mathbf{c}_2$.

(d)
```r
# we can treat the bump functions independently b)
grad <- function(x) {
  if((x - c1) %*% (x - c1) < 1){
    return(c(-4 * c(1 - (x - c1) %*% (x - c1)) * (x - c1)))
  }else if((x - c2) %*% S2_inv %*% (x - c2) < 1){
    return(c(4 * c(1 - (x - c2) %*% S2_inv %*%(x - c2)) * (x - c2) %*% S2_inv))
  }else{
    return(c(0, 0))
  }
}

alpha = 0.15

x0 = c(-0.45, 0.5)
x1 = x0 - alpha * grad(x0)
x2 = x1 - alpha * grad(x1)

print(x1)

## [1] -0.365175  0.421700

print(x2)

## [1] -0.365175  0.421700

print(grad(x1))

## [1] 0 0
```

We can not make any further progress with GD since the gradient is exactly zero.

(e) Start with $\mathbf{x}^{[0]} = (-0.45, 5)^\top$.
Since $\|\mathbf{c}_1 - \mathbf{x}^{[0]}\|_2^2 = 0.5525 < 1$ we know that $\nabla f(\mathbf{x}^{[0]}) = -4(1 - \|\mathbf{x} - \mathbf{c}_1\|_2^2) \cdot (\mathbf{x} - \mathbf{c}_1)^\top = (-0.5655, 0.5220)$.
$\mathbf{x}^{[1]} = \mathbf{x}^{[0]} - 0.15 * (-0.5655, 0.5220)^\top = (-0.3652, 0.422)^\top$.
Since $\|\mathbf{c}_1 - \mathbf{x}^{[1]}\|_2^2 = 1.0001 > 1$ and $\|\mathbf{c}_2 - \mathbf{x}^{[1]}\|_{S_2}^2 = 1.4323 > 1$ the gradient of f is zero at $\mathbf{x}^{[1]}$.
$\Rightarrow \mathbf{x}[2] = \mathbf{x}[1]$

(f)
```r
alpha = 0.15

v = c(0.4, -0.4)
phi = 0.5
x = c(-0.45, 0.5)

xs = x
for (i in 1:15){
  v = phi * v - alpha*grad(x)
  x = x + v
  xs = rbind(xs, x)
}

cont_plot +
  geom_line(data = as.data.frame(xs), aes(x=V1, y=V2), color="red")
```