1) Consider task of optimally clustering the following 12 vectors in $\mathbb{R}^2$

X											
0.7	0.8	0.3	0.7	0.7	0.2	0.1	0.5	1.0	0.3	0.6	0.2
0.8	0.3	0.5	0.7	0.9	1.0	0.5	0.1	0.1	0.3	0.8	0.3

into four distinct clusters under the L_2 distance. Write a `Matlab` code that heuristically solves the clustering problem using the k-means algorithm and succinctly describe it (the fundamental steps, leaving aside the unnecessary details). Then run it on the data above; with the 12 vectors numbered 1, 2, ..., 12 from left to right, use the clusters $\{1, 2, 3\}$, $\{4, 5, 6\}$, $\{7, 8, 9\}$, $\{10, 11, 12\}$ as the initial solution. Comment on the behaviour of the algorithm and report the obtained solution (the centroids and the clusters) and its cost. Optionally, discuss how you could determine if the obtained solution is optimal, and possibly do so.

SOLUTION

The clustering problem in the L_2 norm requires finding four vectors $c_p \in \mathbb{R}^2$, $p = 1, \ldots, 4$ that solve

$$\min\{ f(c) = \textstyle\sum_{1,\ldots,12} \min_{1,\ldots,4} \| c_p - X_i \|_2^2 \,:\, c \in \mathbb{R}^{4\times 2} \}$$

or equivalently, with the introduction of binary variables,

$$\begin{array}{lr} \min \sum_{i=1,\ldots,12} \sum_{p=1,\ldots,4} z_{ip} \| c_p - X_i \|_2^2 & \\ \sum_{p=1,\ldots,4} z_{ip} = 1 & i = 1, \ldots, 12 \\ z_{ip}\{0, 1\} & p = 1, \ldots, 4 \,,\, i = 1, \ldots, 12 \end{array}$$

A way to implement it is to define, together with $X \in \mathbb{R}^{12\times 2}$ and $c \in \mathbb{R}^{4\times 2}$, a vector $k \in \{1, \ldots, 4\}^{12}$ indicating the cluster number, i.e., $k(i) = p$ meaning that X_i belongs to the cluster of centroid c_p. The required starting point corresponds to

$$k = [1, 1, 1, 2, 2, 2, 3, 3, 3, 4, 4, 4]$$

The algorithm then just iterates between forming the optimal centroids corresponding to the given clusters in k by just computing their mean, e.g.,

```
for p = 1 : 4
    c( p , : ) = mean( X( clusters == p , : ) , 1 );
 end
```

(note the importance of the ", 1" parameter, without which the when the cluster is a singleton the result is a single number: the mean of the entries), and then recomputing the clusters as the points having the minumum distance from the given centroids, e.g.,

```
for i = 1 : 12
    dist = vecnorm( c - X( i , : ) , 2 , 2 );
    [ ~ , ci ] = min( dist );
    clusters( i ) = ci;
end
```

The process ends when the objective value, computed as in

```
v = sum( vecnorm( X - c( clusters , : ) , 2 , 2 ) .^2 );
```

stops strictly decreasing from one iteration to the next.

Applied to the given data the algorithm should perform 4 iterations, with the objective value starting at 1.2467 and terminating at 0.24833, with the final centroids

c			
0.6750	0.2000	0.7667	0.2250
0.8000	1.0000	0.1667	0.4000

and the corresponding final clusters

$$k = [1, 3, 4, 1, 1, 2, 4, 3, 3, 4, 1, 4]$$

(note how cluster 2 is in fact a singleton corresponding to $X_6 = [0.1, 1.0]$).

To verify whether the solution is optimal one could write an exact MIQP formulation of the problem, such as

$$
\begin{array}{lll}
\min \ \sum_{i=1,\dots,12} \sum_{p=1,\dots,4} \| v_{ip} \|_2^2 & & \\
\quad (\overline{x} - X_i) z_{ip} \geq v_{ip} \geq (\underline{x} - X_i) z_{ip} & p = 1, \dots, 4 \ , & i = 1, \dots, 12 \\
\quad c_p - X_i z_{ip} - \underline{x}(1 - z_{ip}) \geq v_{ip} \geq c_p - X_i z_{ip} - \overline{x}(1 - z_{ip}) & p = 1, \dots, 4 \ , & i = 1, \dots, 12 \\
\quad \overline{x} \geq c_p \geq \underline{x} & p = 1, \dots, 4 & \\
\quad \sum_{p=1,\dots,4} z_{ip} = 1 & & i = 1, \dots, 12 \\
\quad z_{ip} \in \{ 0 \, , \, 1 \} & p = 1, \dots, 4 \ , & i = 1, \dots, 12
\end{array}
$$

for properly defined worst-case bounds $\overline{x}$ and $\underline{x}$ (the maximum and minimum of X over the columns, respectively) and then solving it with an exact MIQP solver. Doing so would reveal that the optimal solution has value 0.24833, i.e., exactly the one obtained by the k-means heuristic, which in this case is therefore exact.

2) Let f be a C^2 function. Define the concepts of L-smoothness and τ-convexity. Then state and prove what is the optimal fixed stepsize for the gradient method for an L-smooth and τ-convex function, and what is the corresponding convergence rate.

SOLUTION

f is L-smooth if the gradient of f is Lipschitz continuous, i.e.,

$$|| \nabla f(x) - \nabla f(y) || \leq L|| x - y || \qquad \forall x , y$$

f is τ-convex if it is strongly convex modulus τ, i.e., $f(x) - \tau/2 \| x \|^2$ is convex $\forall x$, or alternatively $\alpha f(x) + (1-\alpha) f(z) \geq f(\alpha x + (1-\alpha)z) + \tau/2\alpha(1-\alpha)\| z - x \|^2$ for all x, z and $\alpha \in [0, 1]$. For a convex C^2 function the two conditions are also equivalent to

$$\tau I \preceq \nabla^2 f(x) \preceq LI \equiv \tau \leq \lambda^n \leq \lambda^1 \leq L \qquad \forall x$$

where as usual λ^1 and λ^n are the maximum and minimum eigenvalue of $\nabla^2 f(x)$, respectively.
The generic iterate of the gradient method with fixed stepsize reads

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

Owing to the fact that $\nabla f(x_*) = 0$, we can then write

$$x^{k+1} - x_* = x^k - x_* - \alpha(\nabla f(x^k) - \nabla f(x_*))$$

Since $f \in C^2$, ∇f is in particular continuous and we can apply the Mean Value Theorem on ∇f to establish that

$$\exists\, w \in [x^k , x_*] \text{ s.t. } \nabla f(x^*) - \nabla f(x_*) = \nabla f^2(w)(x^k - x_*)$$

This finally yields

$$x^{k+1} - x_* = x^k - x_* - \alpha \nabla f^2(w)(x^k - x_*) = (I - \alpha \nabla f^2(w))(x^k - x_*)$$

whence,

$$\| x^{k+1} - x_* \| \leq \| I - \alpha \nabla f^2(w) \| \| x^k - x_* \|$$

If for some α one has $r = \| I - \alpha \nabla f^2(w) \| < 1$ then the algorithm is linearly convergent with rate r. In particular, then, the optimal stepsize will be the one solving

$$\min\{ \| I - \alpha \nabla f^2(w) \| \; : \; \alpha \geq 0 \}$$

Now, by definition of the L_2 norm for matrices,

$$\| I - \alpha \nabla f^2(w) \| = \max\{ | 1 - \alpha \lambda_1(\nabla f^2(w)) | \, , \, | 1 - \alpha \lambda_n(\nabla f^2(w)) | \}$$

It is easy to check that when $1 - \alpha\lambda_n \geq 1 - \alpha\lambda_1 \geq 0$, increasing α decreases the max. Symmetrically, when $0 \leq \alpha\lambda_n - 1 \leq \alpha\lambda_1 - 1$, decreasing α decreases the max. Thus, the optimal α must be s.t. $1 - \alpha\lambda_n > 0$ and $1 - \alpha\lambda_1 < 0$, i.e., $r = \max\{ -1 + \alpha\lambda_1 \, , \, 1 - \alpha\lambda_n \}$. Of course λ_1 and λ_n depend on w and are unknown in general, but $L \geq \lambda_1$ and $\tau \leq \lambda_n$ whence $r \leq \bar{r} = \max\{ -1 + \alpha L \, , \, 1 - \alpha\tau \}$. Now, clearly the two terms in the max behave symmetrically: if one grows the other decreases and vice-versa. It is therefore obvious that the optimal α is the one where they are equal (for otherwise one could decrease the larger one and increase the smaller one). It is easy to check that this happens when $\alpha = 2 / (L + \tau)$. In fact,

$$-1 + 2L / (L + \tau) = (-L - \tau + 2L) / (L + \tau) = (L - \tau) / (L + \tau)$$

$$1 - 2\tau / (L + \tau) = (L + \tau - 2\tau) / (L + \tau) = (L - \tau) / (L + \tau)$$

All in all, with $\alpha = 2 / (L + \tau)$ one has $\| x^{k+1} - x_* \| \leq r^k \| x^1 - x_* \|$ with $r = (L - \tau) / (L + \tau) < 1$, i.e., the algorithm converges linearly. An alternative way of writing the result is

$$\bar{r} = (L - \tau) / (L + \tau) = (\bar{\kappa} - 1) / (\bar{\kappa} + 1)$$

with with $\bar{\kappa} = L / \tau$ the worst-case condition number of ∇f^2. This yields the intuitive result that the convergence rate worsens the more ∇f^2 can be ill-conditioned.

3) Consider the following multiobjective optimization problem:

$$\begin{cases} \min\ (3x_1^2 + x_2^2 - x_1x_2,\ 2x_1 - x_2) \\ -2x_1 + x_2 - 2 \le 0 \end{cases}$$

(a) Prove that the problem admits a Pareto minimum point.

(b) Find the set of all weak Pareto minima.

(c) Find a suitable subset of Pareto minima.

SOLUTION

(a) Since the feasible set X is nonempty, closed and the objective function f_1 is strongly convex then the problem admits at least a (Pareto) minimum point.

(b) We preliminarly observe that the problem is convex, since the objective and the constraint functions are convex. Therefore the set of weak minima coincides with the set of solutions of the scalarized problems (P_{α_1}), where $0 \le \alpha_1 \le 1$, i.e.

$$\begin{cases} \min\ \alpha_1(3x_1^2 + x_2^2 - x_1x_2) + (1-\alpha_1)(2x_1 - x_2) =: \psi_{\alpha_1}(x) \\ -2x_1 + x_2 - 2 \le 0 \end{cases}$$

For $0 < \alpha_1 \le 1$, ψ_{α_1} is strongly convex so that P_{α_1} admits a unique optimal solution which is a minimum of (P).

We note that (P_{α_1}) is convex and differentiable and ACQ holds at any $x \in X$; then the KKT system provides a necessary and sufficient condition for an optimal solution of (P_{α_1}). KKT system is given by:

$$\begin{cases} 6\alpha_1x_1 - \alpha_1x_2 + 2 - 2\alpha_1 - 2\lambda = 0 \\ 2\alpha_1x_2 - \alpha_1x_1 - 1 + \alpha_1 + \lambda = 0 \\ \lambda(-2x_1 + x_2 - 2) = 0 \\ 0 \le \alpha_1 \le 1,\ \lambda \ge 0 \\ -2x_1 + x_2 - 2 \le 0 \end{cases}$$

By eliminating λ in the first equality, we obtain:

$$\begin{cases} \alpha_1(4x_1 + 3x_2) = 0 \\ \lambda = -2\alpha_1x_2 + \alpha_1x_1 + 1 - \alpha_1 \\ \lambda(-2x_1 + x_2 - 2) = 0 \\ 0 \le \alpha_1 \le 1,\ \lambda \ge 0 \\ -2x_1 + x_2 - 2 \le 0 \end{cases} \tag{1}$$

For $\alpha_1 = 0$, we obtain $\lambda = 1$, $-2x_1 + x_2 - 2 = 0$, so that

$$\text{Weak Min(P)} \supseteq \{(x_1, x_2) : -2x_1 + x_2 - 2 = 0\}.$$

For $0 < \alpha_1 \le 1$ the system (1) becomes

$$\begin{cases} 4x_1 + 3x_2 = 0 \\ \lambda = -2\alpha_1x_2 + \alpha_1x_1 + 1 - \alpha_1 \\ \lambda(-2x_1 + x_2 - 2) = 0 \\ 0 < \alpha_1 \le 1,\ \lambda \ge 0 \\ -2x_1 + x_2 - 2 \le 0 \end{cases} \tag{2}$$

From the complementarity condition $\lambda(-2x_1+x_2-2) = 0$, we have the two cases: $I)$ $\lambda = 0$, $II)$ $\lambda \ne 0$ $(or, -2x_1+x_2-2 = 0)$. In case $I)$, $\lambda = 0$, by the first two equations we obtain:

$$x_1 = \frac{3}{11}\left(1 - \frac{1}{\alpha_1}\right),\ x_2 = -\frac{4}{11}\left(1 - \frac{1}{\alpha_1}\right)$$

Imposing the feasibility condition given by the last inequality, and recalling that $0 < \alpha_1 \le 1$, we obtain $\frac{5}{16} \le \alpha_1 \le 1$ so that

$$Min(P) \supseteq \{(x_1, x_2) : x_2 = -\frac{4}{3}x_1, -\frac{3}{5} \le x_1 \le 0\}.$$

In case II), $\lambda \neq 0$ the the system (2) becomes

$$\begin{cases} 4x_1 + 3x_2 = 0 \\ \lambda = -2\alpha_1 x_2 + \alpha_1 x_1 + 1 - \alpha_1 \\ -2x_1 + x_2 - 2 = 0 \\ 0 < \alpha_1 \leq 1, \ \lambda > 0 \end{cases} \tag{3}$$

and we obtain the unique solution $(x_1, x_2) = (-\frac{3}{5}, \frac{4}{5})$, a point that we had already found as a minimum.

In conclusion:

$$\text{Weak Min(P)} = \{(x_1, x_2) : -2x_1 + x_2 - 2 = 0\} \quad \cup \{(x_1, x_2) : x_2 = -\frac{4}{3}x_1, -\frac{3}{5} \leq x_1 \leq 0\}.$$

$$Min(P) \supseteq \{(x_1, x_2) : x_2 = -\frac{4}{3}x_1, -\frac{3}{5} \leq x_1 \leq 0\}.$$

It is possible to show that $Min(P)$ actually coincides with the found minima.

4) Consider the following matrix game:

$$C = \begin{pmatrix} 5 & 4 & 3 & 5 \\ 6 & 7 & 8 & 2 \\ 5 & 3 & 4 & 4 \end{pmatrix}$$

(a) Find the set of pure strategies Nash equilibria, if any. Alternatively, show that no pure strategy Nash equilibrium exists.

(b) Find a mixed strategy Nash equilibrium which is not a pure strategy Nash equilibrium, if any. Alternatively, show that no Nash equilibrium of such kind exists.

SOLUTION (a) Considering Player 1, the possible couples of pure strategies Nash equilibria could be $(1,1)$, $(3,1)$, $(3,2)$, $(1,3)$ and $(2,4)$ (minimal components on the columns), while considering Player 2 the pure strategies Nash equilibria could be $(1,1)$, $(1,4)$,$(2,3)$ and $(3,1)$ (maximal components on the rows). The common couples $(1,1)$ and $(3,1)$ are pure strategies Nash equilibria.

(b) Consider the linear optimization problem associated with Player 1:

$$\begin{cases} \min \quad v \\ 5x_1 + 6x_2 + 5x_3 \le v \\ 4x_1 + 7x_2 + 3x_3 \le v \\ 3x_1 + 8x_2 + 4x_3 \le v \\ 5x_1 + 2x_2 + 4x_3 \le v \\ x_1 + x_2 + x_3 = 1 \\ x \ge 0 \end{cases} \tag{4}$$

Pure strategies for Player 1 correspond to the solutions $(x^1, v^1) = (1,0,0,5)$ and $(x^2, v^2) = (0,0,1,5)$ and $y = (1,0,0,0,5)$ is a dual solution of (4) associated with the pure strategy $(1,0,0,0)$ of Player 2. Since the problem is linear then any convex combination

$$\alpha(x^1, v^1) + (1-\alpha)(x^2, v^2), \quad \alpha \in [0,1]$$

is an optimal solution of (4).

For example, for $\alpha = \frac{1}{2}$ we obtain the solution $\hat{x} = (\frac{1}{2}, 0, \frac{1}{2}, 5)$ so that

$$\hat{x} = (\frac{1}{2}, 0, \frac{1}{2}), \quad \hat{y} = (1,0,0,0)$$

is a mixed-strategies Nash equilibrium.