Introduction to Minimax

prepared by Leiming Qian

based on the book with the same title by V. F.Dem'yanov and V. N. Malozemov

Motivation

- Minimax is an important principle in optimal selection of parameters.

- In some cases minimax solution has the remarkable property known as alternation.

- Some practical examples:
  - best polynomial approximation.
  - Mandel'shtam's Problem in circuit theory.
  - matrix games.

General Form of Minimax Problem

- Ω is a convex closed subset of E_n, G is a bounded closed subset of E_m, we want to solve for X^*:

$$\max_{Y \in G} F(X^*, Y) = \min_{X \in \Omega} \max_{Y \in G} F(X, Y) \tag{1}$$

- If $F(X, Y)$ is linear, then problem 1 is known as *linear minimax problem*; if $\Omega \neq E_n$, it is known as *constrained minimax problem*.

- X is known as the *parameter*, $F(X, Y)$ is known as the *cost function*.

Three Basic Ideas to Solve Minimax

- Search for *extremal basis* G_r:

$$\max_{Y \in G_r} F(X^*, Y) = \min_{X \in \Omega} \max_{Y \in G} F(X, Y)$$

- Minimization of the *maximum function* $\phi(X)$:

$$\phi(X) = \max_{Y \in G} F(X, Y)$$

- Determination of a *saddle point* (X^*, Y^*):

$$F(X^*, Y) \le F(X^*, Y^*) \le F(X, Y^*)$$

Outline of the Introduction

- Problem of best polynomial approximation.
- Linear minimax problem without constraints.
- Linear minimax problem with constraints.

Best Polynomial Approximation

- Suppose we know $t_0 < t_1 < t_2 < \ldots < t_N$, and corresponding values $y_k = y(t_k)$.

- The goodness of fit of an algebraic polynomial $P_n(A,t)$ of degree at most $n (n \leq N)$,

$$
\begin{aligned}
P_n(A,t) &= a_0 + a_1 t + \ldots + a_n t_n, \qquad A = (a_0, \ldots, a_n) \\
\rho &= \inf_A \max_{k \in |0:N|} |y_k - P_n(A, t_k)|
\end{aligned}
$$

- The polynomial $P_n(A^*, t)$ for which ρ is minimal is known as the best approximation polynomial.

Order of Approximation n

- The relationship between n and N is essential.

- If $n = N$, the solution is the interpolating polynomial, $\rho = 0$.

- If $n = N - 1$, this is known as Chebyshev interpolation, the fundamental case.

- Chebyshev interpolation provides the best approximation in the general case of $N > n + 1$.

Chebyshev Interpolation $N = n + 1$

- Setting $N = n + 1$, we have:

  Theorem 1 *There exists a unique polynomial of best approximation.* $P_n(A^*, t)$ *is this polynomial if and only if, for some* h,

  $$(-1)^k h + P_n(A^*, t_k) = y_k, \quad k \in [0 : n+1]$$

  in that case, $\rho = |h|$.

- The construction of a polynomial satisfying Theorem 1 is known as Chebyshev Interpolation.

Chebyshev Polynomial Construction

- h can be explicitly expressed as:

$$\alpha_k = \frac{\frac{(-1)^k}{(t_k - t_0)\ldots(\ )\ldots(t_k - t_{n+1})}}{\sum_{\nu=0}^{n+1} \frac{(-1)^k}{(t_\nu - t_0)\ldots(\ )\ldots(t_\nu - t_{n+1})}} \tag{2}$$

$$h = \sum_{k=0}^{n+1} (-1)^k \alpha_k y_k \tag{3}$$

$$k \in [0 : n+1]$$

- The expression for h can be greatly simplified under specific choice of $\{t_k\}$.

General Discrete Case $N > n + 1$

- We have the following theorem:

  Theorem 2 *There exists a unique polynomial of best approximation. $P_n(A^*, t)$ is this polynomial if and only if it is a Chebyshev interpolation polynomial for some extremal basis σ^*.*

- The exact solution can be obtaiced by carrying out finitely many Chebyshev interpolations (de la Vallee-Poussin algorithm).

de la Vallee-Poussin Algorithm

- Construct a basis $\sigma_i = \{t_{i_0} < t_{i_1} < \ldots < t_{i_{n+1}}\}$, the initial one is chosen arbitrarily.
- Carry out Chebyshev interpolation and construct $P_n(A(\sigma_i), t)$.
- Construct a new basis σ_{i+1} by modifying one point in σ_i, such that $\rho(\sigma_i) < \rho(\sigma_{i+1})$.
- Go back to step 1 until solution reached.
- This procedure is finite because the number of valid bases is finite.

R-Algorithm

- de la Vallee-Poussin algorithm utilizes previous computation poorly (inefficient).
- R-algorithm replaces several points from the original basis and leads to a higher increase of ρ.
- The initial bcsis σ_0 in the R-algorithm must be chosen such that $\rho(\sigma_0) > 0$.
- R-algorithm also reaches the solution in finite number of steps.

The Discrete Minimax Problem

- Let $f_i(X), i \in [0 : N], X = (x_1, x_2, \ldots, x_n)$ be functions defined on euclidean n-space E_n. We define

$$\phi(X) = \max_{i \in [0:N]} f_i(X)$$

- We want to find point X^* at which $\phi(X)$ is minimized, $\phi(X)$ is known as the *maximum function*.

Important Properties of $\phi(X)$

- Let $\Omega \subset E_n$ be a convex set, if all $f_i(X)$ are convex on Ω, then $\phi(X)$ is also convex on Ω.
- Let $f_i(X)$ be continuous fsnctions with continuous derivatives of order up to l in some neighborhood $S_\sigma(X_0)$ of X_0, then $\phi(X)$ admits the following presentation near X_0 in direction g, $\|g\| = 1$:

$$\begin{aligned}
\phi(X_0 + \alpha g) &= \phi(X_0) + \sum_{k=l}^{l} \frac{\alpha^k}{k!} \frac{\partial^k \phi(X_0)}{\partial g^k} + o(g; \alpha^l) \\
\frac{\partial^k \phi(X_0)}{\partial g^k} &= \max_{i \in R_k(X_0, g)} \frac{\partial^k f_i(X_0)}{\partial g^k}
\end{aligned}$$

Necessary Condition for a Minimax

- If $f_i(X)$ are continuously qifferentiable on E_n:

  Theorem 3 $\phi(X)$ *has a minimum at* X^**, it is necessary, and if* $\phi(X)$ *is convex, also sufficient that*

$$\inf_{\|g\|=1} \max_{i \in R(X^*)} \left(\frac{\partial f_i(X^*)}{\partial X},\ g \right) \geq 0$$

  or, equivalently,

$$\inf_{\|g\|=1} \frac{\partial \phi(X^*)}{\partial g} \geq 0$$

Geometric Interpretation

- For any fixed $X \in E_n$, we consider the set:

$$H(X) = \left\{ Z \in E_n | Z = \frac{\partial f_i(X)}{\partial X}, i \in R(X) \right\}$$

- Let $L(X)$ denote the convex hull of $H(X)$:

$$L(X) = \left\{ Z = \sum_{i \in R(X)} \alpha_i \frac{\partial f_i(X)}{\partial X} \,\Big|\, \alpha_i \geq 0, \sum_{i \in R(X)} \alpha_i = 1 \right\}$$

- We have the following equivalency theorem:

Theorem 4 *Theorem 3 is equivalent to the inclusion*

$$0 \in L(X^*)$$

Sufficient Condition for a Local Minimax

- If $f_i(X)$ are continuously differentiable on E_n:

  Theorem 5

  $$\psi(X) = \min_{\|g\|=1} \max_{i \in R(X)} \left(\frac{\partial f_i(X)}{\partial X}, g \right)$$

  if $\psi(X^*) = r > 0$, then X^* is a strict local minimum point for $\phi(X)$.

- Similarly, we have second-order sufficient conditions.

Estimates for $\mu = \min \phi(X)$

- If $f_i(X)$ are continuously differentiable on E_n:

  Theorem 6 *Assume for $X_0 \in E_n$, there is a set of indices $Q \subset [0 : N]$ such that*

  $$\min_{\|g\|=1} \max_{i \in Q} \left(\frac{\partial f_i(X)}{\partial X}, g \right) \geq 0$$

  if $f_i(X)$ are convex for all $i \in Q$, then

  $$\min_{i \in Q} f_i(X_0) \leq \mu \leq \phi(X_0)$$

- Similarly, we have a more precise second-order estimate.

Method of Coordinatewise Descent

- Select initial approximation X_0 arbitrarily, assume set $M(X_0) = \{X \mid \phi(X) \le \phi(X_0)\}$ is bounded.

- Suppose kth approximation determined, to construct X_{k+1}, we consider the straight line

$$X = X_{k1}(\alpha) = X_{k1} + \alpha e_1, \quad -\infty < \alpha < \infty$$

  and determine α_{k1} such that

$$\phi(X_{k1}(\alpha_{k1})) = \min_{\alpha \in (-\infty, \infty)} \phi(X_{k1}(\alpha))$$

- Perform the same for e_2, etc. The limit might NOT always be a local minimum point!

Method of Steepest Descent

- Based on the fact that $\phi(X)$ has a unique direction of steepest descent $g(\bar{X})$ at $\bar{X}$.
- Suppose kth approximation determined, if $\psi(X_k) \leq 0$, we construct the vector $g_k = g(X_k)$

  $$X = X_k(\alpha) = X_k + \alpha \ g_k, \quad \alpha \geq 0$$

  and find α_k for which

  $$\phi(X_k(\alpha_k)) = \min_{\alpha \in [0,\infty)} \phi(X_k(\alpha))$$

- More flexible, but also does not guarantee convergence to a local minimum point.

First Method of Successive Approximations

- We make the following definitions:

$$\begin{aligned} R_\epsilon(X) &= \{i \mid \phi(X) - f_i(X) \le \epsilon\}, \quad \epsilon \ge 0 \\ \psi_\epsilon(X) &= \min_{\|g\|=1} \max_{i \in R_\epsilon(X)} \left(\frac{\partial f_i(X)}{\partial X},\ g \right) \end{aligned}$$

- We also define ϵ-stationary point X_ϵ^* and ϵ-steepest descent direction $g_\epsilon(X)$.

First Method of Successive Approximations

- We fix two parameters $\epsilon_0 > 0$ and $\rho_0 > 0$, take initial approximation X_0:
- Suppose we found k-th approximation X_k, if $\psi(X_k) < 0$, then we check sequence $\epsilon_\nu = \epsilon_0/2^\nu$ and find the first ν such that

$$\psi_{\epsilon_\nu}(X_k) \leq -\frac{\rho_0}{\epsilon_0}\epsilon_\nu$$

- Set $g_k = g_{\epsilon_k}(X_k)$, the ϵ-steepest descent direction and update X_k.
- This algorithm guarantees convergence to local minimum if $\phi(X)$ is convex.

More Methods of Successive Approximations

- Second method of successive approximations: ϵ-algorithm. More advanced techniques in updating ϵ, faster convergence.

- Third method of successive approximations: D-function.

$$D(X) = \inf_{\epsilon \in [0,\bar{\epsilon}]} \epsilon \psi_\epsilon(X), \quad \bar{\epsilon} > 0$$

D-function is always continuous while $\phi(X)$ is not. We operate in the D-function space.

Summary of Successive Approximations

- The first method is mainly of theoretical interest as direct generalization of steepest descent.

- The second method is the most convenient for practical use.

- The third method is highly effective for small n and N.

- The most important component is determination of ϵ-stationary points.

Constrained Discrete Minimax Problem

- Let $f_i(X), i \in [0 : N], X = (x_1, x_2, \ldots, x_n)$ be functions defined on $\Omega \subset E_n$. We define

$$\phi(X) = \max_{i \in [0:N]} f_i(X)$$

- We want to find point $X^* \in \Omega$ at which $\phi(X)$ is minimized (Ω is not necessarily bounded.)
- Most theorems and methods for unconstrained problem still apply after slight modifications.
- Usually linearize the constraints and apply linear programming techniques, otherwise non-trivial.