CÁC QUÁ TRÌNH TỰ THÍCH NGHI TRONG PHƯƠNG PHÁP DÒ TÌM NGẪU NHIÊN VÀ ỨNG DỤNG CỦA NÓ TRONG CÁC BÀI TOÁN THỦY VĂN

PTS. Nguyễn Hữu Bảo
Trường Đại học Thủy lợi, Hà Nội

1. Đặt vấn đề

Trong các mô hình tính toán thủy văn (như dự báo, điều khiển v.v.) thường chứa nhiều tham số và việc xác định các tham số này tùy thuộc vào việc thỏa mãn một số điều kiện ràng buộc nào đó và phải tối ưu hóa một hàm mục tiêu nào đó. Các tham số đó được gọi là các tham số điều khiển và bài toán xác định các biến điều khiển để tối ưu một hàm mục tiêu cho trước với những ràng buộc cho trước là một bài toán tối ưu thông thường mà không ai xa lạ trong lĩnh vực ứng dụng thủy văn của nó. Tuy nhiên, nếu kể đến những nhiễu loạn ngẫu nhiên và nếu hàm mục tiêu cũng là một hàm ngẫu nhiên, ví dụ nó chứa đựng các quan sát đo đạc (hay nói cách khác là các thể hiện của một quá trình ngẫu nhiên nào đó) thì bài toán trở nên không đơn giản.

Ví dụ, bài toán xác định mô hình truyền nước trên một đoạn sông, nếu sử dụng công thức tích phân Duemen [1] thì công thức để xác định dòng chảy ra phụ thuộc vào 2 tham số τ và n chưa biết và cần xác định nó thỏa mãn điều kiện ràng buộc nào đó sao cho sai số giữa kết quả tính toán lưu lượng nước chảy ra với các quan sát thực đo ở cùng một số thời điểm trong quá khứ là nhỏ nhất.

Một ví dụ khác là việc xác định hàm mật độ phân phối xác suất Q_{max} cho lưu lượng nước tại một trạm đo nào đó. Nếu giả định nó có phân phối loga chuẩn với 2 tham số μ và σ^2 chưa biết thì việc ước lượng 2 tham số này như thế nào để hàm mật độ tính toán phải sát nhất với mật độ theo thực đo là một việc quan trọng và đầu tiên cho các nghiên cứu tiếp theo. Có thể lấy ra hàng loạt các ví dụ khác như việc xác định lượng nước xả và dùng cho các hồ chứa như thế nào để tối ưu hóa hàm mục tiêu điện năng, tưới tiêu v.v. mà vẫn đảm bảo các điều kiện ràng buộc về an toàn hồ chứa, an toàn môi trường và khống chế thiệt hại hoa màu, v.v.

Tất cả các bài toán này đều có thể đưa về việc phát biểu bằng công cụ toán học bài toán tối ưu ngẫu nhiên sau đây:

2. Bài toán tối ưu ngẫu nhiên

Xét không gian tham số $X = (X_1, X_2... X_n)$ thuộc R^n là không gian các trạng thái của biến điều khiển $(x_1, x_2... x_n)$. Gọi S là tập con của X gồm các trạng thái của biến điều khiển $(x_1... x_n)$ sao cho nó thỏa mãn 1 số điều kiện nào đó và ta gọi là miền ràng buộc. Gọi $H = H(X, E)$ là một hàm mục tiêu xác định trên R^{n+m}, phụ thuộc không chỉ vào các biến điều khiển $x_1, x_2, ..., x_n$ mà còn phụ thuộc vào cả véc tơ ngẫu nhiên $E = (\varepsilon, \varepsilon, ..., \varepsilon_m)$ là tác động ngẫu nhiên nào đó đôi khi biết được mật độ $p(E)$ của nó. Gọi

$$Q(X) = \int H(X, E)\, p(E)\, dE = M_E H(X, E)$$

là một hàm mục tiêu theo kỳ vọng, đó là một hàm tất định.

Bài toán tối ưu ngẫu nhiên là bài toán tìm nghiệm tối ưu $X^* \in S$ sao cho nó cực tiểu hóa phiến hàm giá $Q(X)$ nói trên:

$$Q(X^*) = \min_{X \in S} M_E H(X, E) \qquad (1)$$

Ta xét ví dụ sau:

Trong bài toán xác định lưu lượng nước chảy ra, nếu không kể đến tác động của dòng chảy vào 2 bên sườn mà chỉ quan tâm đến lưu lượng nước $Q_H(t)$ của dòng chảy vào tại thời điểm t thì có thể xác định lưu lượng dòng chảy ra $Q(t)$ theo công thức tích phân Duemen [1]

$$\widehat{Q}(t) = \int_0^t \frac{1}{\tau\Gamma(n)} \left(\frac{t-x}{\tau}\right)^{n-1} e^{-\frac{t-x}{\tau}} Q_H(x)dx \qquad (2)$$

Rõ ràng là $\widehat{Q}(t)$ được xác định chính xác hay không là tùy thuộc vào việc ước lượng 2 tham số τ và n. Không gian tham số X ở đây gồm 2 chiều (τ, n). Miền ràng buộc S có thể lấy là miền chữ nhật.

$$S = \{\tau_a \le \tau \le \tau_b;\ n_a \le n \le n_b\}$$

Hàm giá là sai số trung bình phương giữa lưu lượng $\widehat{Q}(t)$ tính toán và $\overline{Q}(t)$ thực đo ở cùng một thời điểm t trong quá khứ,

$$Q(X) = \sum_{i=1}^{p} \left[\widehat{Q}(t_i) - \overline{Q}(t_i)\right]^2 \qquad (3)$$

nếu ta có p quan sát $\overline{Q}(t_1), \overline{Q}(t_2) ... \overline{Q}(t_p)$ ở quá khứ. Ta cần xác định (τ^*, n^*) là nghiệm tối ưu thỏa mãn điều kiện:

$$(\tau^*, n^*) \in S \text{ và } Q(\tau^*, n^*) = \min_{X \in S} Q(X)$$

Các ví dụ khác đều có thể đưa về bài toán trên một cách tương tự. Vấn đề là: thuật toán để tìm nghiệm tối ưu nói trên.

3. Thuật toán dò tìm ngẫu nhiên

Các bài toán tối ưu dạng (1) và được gọi là bài toán quy hoạch ngẫu nhiên hay điều khiển ngẫu nhiên tùy theo ý nghĩa và cách đặt bài toán. Các bài toán này đã được các tác giả Rastrigin trong [2] và [3] và Ermoliep trong [4] nghiên cứu từ những năm của thập niên 70. Các tác giả đã đưa ra phương pháp dò tìm ngẫu nhiên và Rastrigin trong [2] đã đưa ra các quá trình tự thích nghi

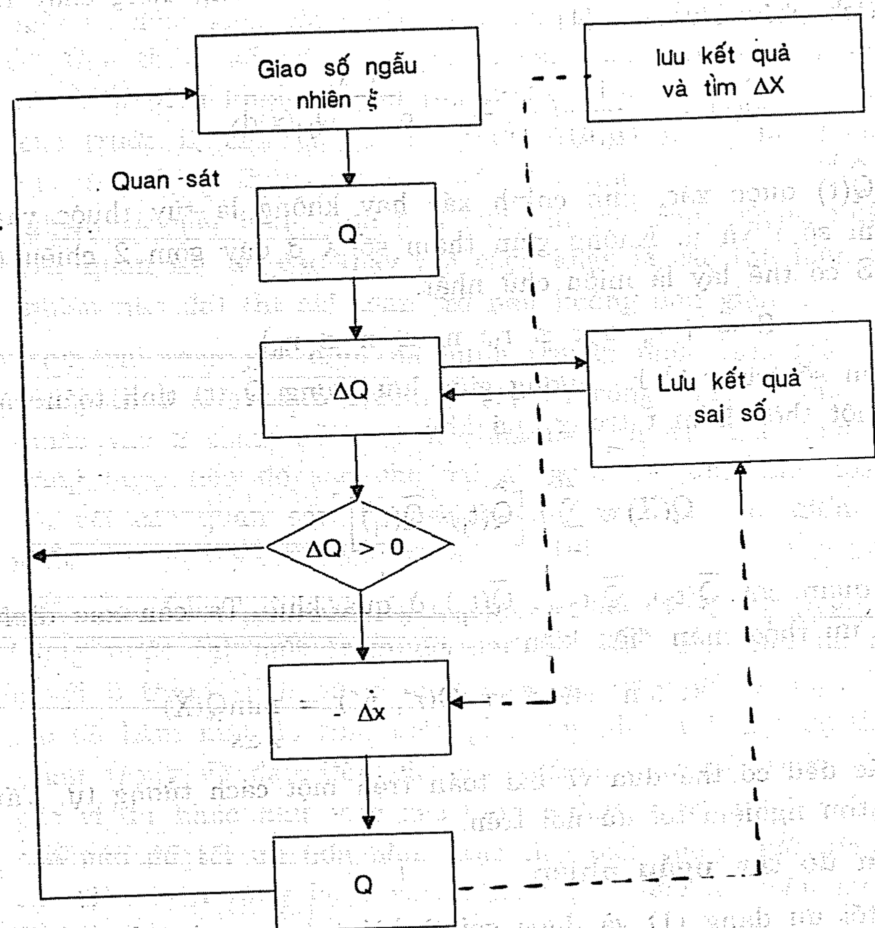
$$X_{i+1} = F(X_i)$$

với F là một toán tử thích nghi được xác định sao cho dãy $X_1, X_2 ...$ sẽ hội tụ (theo xác suất hay theo kỳ vọng) về nghiệm tối ưu X^* theo nghĩa "thích nghi" trong từng bước tính toán. Trong [3] các tác giả đã trình bày phương pháp gradien với quá trình tự thích nghi Robinu - Monro như sau:

$$X_{i+1} = X_i - \Gamma_i \nabla Q(X_i)$$

trong đó $\nabla Q(X_i)$ là toán tử sai phân xấp xỉ cho Grad $Q(X_i)$ còn Γ_i là một là ma trận vuông cần chọn sao cho thỏa mãn 4 điều kiện cần của định lý hội tụ Martingale. Các quá trình này có ý nghĩa ứng dụng hết sức to lớn. Tuy nhiên, việc chọn ma trận Γ_i như thế nào để thỏa mãn điều kiện cần cho định lý hội tụ Martingale là một việc rất khó và tùy thuộc vào cách xây dựng các bài toán. Và lại, việc xác định tham số điều khiển X_0 của bước đầu tiên thì không đề cập đến.

Trong một số bài toán đơn giản thì giải quyết 2 vấn đề trên không gặp nhiều khó khăn. Ví dụ, việc chọn X_0 có thể từ các phương pháp thông thường của thủy văn như chia bước đủ nhỏ miền ràng buộc S và thuật toán để xác định quá trình tự thích nghi X_0, X_1, X_2... có thể tính theo sơ đồ sau:



Với sơ đồ đó, các điều kiện hội tụ Martingale đã được thỏa mãn còn ξ là số ngẫu nhiên có phân phối đều trên đoạn [-1, 1] [2].

4. Ứng dụng vào một mô hình tính toán thủy văn

Trong việc khảo sát một đoạn sông Mekông từ Paksé về Banchannoi [5] chúng tôi đã sử dụng tích phân Duemen để xác định dòng chảy ra và chưa tính đến lượng nhập khu giữa.

Áp dụng phương pháp dò tìm ngẫu nhiên nói trên cùng với quá trình tự thích nghi theo sơ đồ trên, việc tính toán ngay từ bước đầu đã tỏ ra rất có hiệu quả.

Sai số tính toán và thực do giảm hẳn so với các phương pháp của thủy văn hiện sử dụng. Nếu cải tiến mô hình thêm chút ít và trong công thức tích phân Duemen có thêm một số hạng nữa khi xét thêm dòng chảy khu giữa, hy vọng sai số còn giảm hơn nữa. Dù sao đây cũng là bước ứng dụng mang tính chất thử nghiệm để khẳng định một vài ưu điểm của phương pháp này.

TÀI LIỆU THAM KHẢO

1. Н. П. Ченова. "Модель формирования стока на приточном участке реки". Метеорология и гидрология. 1982. No6.

2. Растричин. "Случайный поиск в процессах адаптации" Рига. 1983.

3. Rastrigin L.A. "Contemporary principles to control complex objects" Moscow, 1983.

4. Ермольёв. "Методы решения нелинейных экаремальных задач". Кибернетика, 1966.

5. Nguyễn Quỳnh Hương, Nguyễn Hữu Bảo. Một phương pháp xác định tham số trong mô hình dòng chảy Paksé- Banchannoi. Thông báo Hội nghị khoa học nhân 35 năm thành lập trường Đại học Thủy lợi.