

화학공정 모델을 이용한 데이터 보정 및 gross error의 감지

어수영 · 장태석 · 신동일 · 장광식* · 윤인섭†

서울대학교 응용화학부

*삼성엔지니어링㈜ 공정팀

(1998년 3월 12일 접수, 1999년 5월 20일 채택)

Data Reconciliation and Gross Error Detection Using Chemical Process Models

Sooyoung Eo, Taesuk Chang, Dongil Shin, Gwang Sik Jang* and En Sup Yoon†

School of Chem. Eng., Seoul National University, Seoul 151-742, Korea

*Process Team, Samsung Engineering Co. Ltd., Seoul 135-090, Korea

(Received 12 March 1998; accepted 20 May 1999)

요 약

본 연구에서는 화학공정 측정값들에 대한 데이터 보정 및 gross error 감지를 위한 통합 시스템을 제안하였다. 제안된 시스템은 기본적인 단위 공정 모델 및 물성치 루틴을 갖추고 있어 하나의 독립된 시스템으로 사용이 가능하다. 제안된 데이터 보정 알고리즘은 비선형 방정식계로 표현되는 화학공정의 특성에 적합하도록 구현되었다. 이 방법은 선형 근사로 얻어진 제약식에 대해 반복적으로 계산을 수행하며, 방정식 중실률을 채택하여 계산의 효율성을 높이도록 구성되었다. 또한, gross error 감지를 위해 기존의 측정치 테스트를 개선한 교차점 테스트를 제안하고 사례 연구를 통해 이의 성능을 확인하였다.

Abstract – In this paper an integrated system for data reconciliation and gross error detection of the measurements in chemical processes is suggested. With unit process models and physical property routines, the proposed system itself works as an independent system. The proposed algorithm of data reconciliation is appropriate for the chemical processes that are consisted of the set of nonlinear equations. Since this method takes advantage of the repeated calculations of the constraint equations acquired by linear approximation, it improves the overall efficiency of the calculation. Also, the Point Test algorithm, which is better than the Measurement Test in co-linearity, for gross error detection is also suggested. The performance of the proposed system has been tested through a case study.

Key words: Data Reconciliation, Measurement Error, Gross Error Detection, Model-based

1. 서 론

공정 전반에 걸친 측정치(raw data)에는 개개의 측정 장치의 정확도에 따라 크고 작은 오차가 존재하며, 때로 계기의 고장, 눈금 조정 잘못, 파이프 누출 등의 원인으로 상대적으로 큰 gross error를 포함할 수도 있다. 이러한 오차가 포함된 측정치들은 공장 전체에 대한 정상상태 수치를 만족시키지 못하는 비일관성을 보이며, 데이터 상호 간에도 서로 상반되는 경향을 보이기도 한다. 그리고 기술적 어려움, 경제적 이유 때문에 직접 측정되지 못한 변수들과 각 단위 장치의 성능을 나타내는 매개 변수 등을 추정해야 할 필요가 있다. 이와 같은 데이터들의 처리 수단을 제공하는 것이 데이터 보정(Data Reconciliation) 기법이다. 데이터 보정은 측정 데이터에 존재하는 오차를 제거 혹은 감소시켜 더욱 정확하고 신뢰도 높은 데이터를 얻기 위한 통계적 절차이다.

지금까지의 데이터 보정에 관한 연구는 대부분 물질 수지만을 고려

하여 수행되었으며[1-4], 때로 공정 모델에 에너지 수치를 추가하기도 하였다. 그 이유는 해를 구하기 쉽고, 선형 시스템에 관하여는 통계적 기법의 이론이 잘 적용되기 때문이다. 그러나, 선형 모델만으로는 얻을 수 있는 정보가 상당히 제한되고, 복잡한 공장의 특성을 효과적으로 표현하기 어렵다.

공정 모델을 나타내는 방정식들과 공정 측정치들을 통계적으로 처리함으로써 원하는 정도의 정보를 추출해 낼 수 있다. 이를 통해 한편으로는 공정 모델을 실제 공장과 일치하도록 개선해 나갈 수 있으며, 한편으로는 공정 모델을 측정 상의 문제점을 파악하는 기준으로 사용할 수 있다. 이러한 두 작업, 즉 공정의 특성을 최대한 반영하는 모델 구축과 측정 오차 감지의 수행을 위해서는 상세한 공정 모델(rigorous process model)이 필요하다[5-7].

1-1. 비선형 시스템의 데이터 보정

비선형 시스템에 대한 통계적 기법을 처음 화학공정에 도입한 것은 1973년 Britt와 Luecke[8]에 의해서였다. 이 경우에는 공정의 제약 조건이 일반적인 비선형 방정식으로 표현되므로 기본적으로 선형 근사

†E-mail : esyoon@pslab.snu.ac.kr

를 이용하며, 따라서 반복계산이 필요하다. 이 문제의 해를 구하는데 있어 미측정 변수 혹은 매개 변수에 대한 최대 확률 추정치(maximum likelihood estimates)를 얻기 위해서는 이에 관한 추정을 측정 변수의 보정과 동시에 수행해야 한다. 또한 이로부터 구성되는 Lagrangian에 대한 기준 방정식(normal equation)들이 비선형 방정식이 되므로, Britt와 Luecke는 이를 부분적으로 선형화하여 반복적으로 계산을 수행하였다. 이후 Knepper와 Gorman[9], MacDonald와 Howat[10], Veverka[11] 등에 의해 유사한 방법들이 제안되었다.

1-2. 공정 모델을 이용한 데이터 보정

공정의 정상상태 모델에 근거한 데이터 보정에 관한 연구는 문헌에 보고된 것이 많지 않다. 이 중 Stephenson과 Shewchuck[12]은 기존의 방정식 중심 기법의 모사 프로그램에 데이터 보정 기능을 추가하여 공정의 모사와 데이터 보정을 동시에 수행할 수 있는 프로그램을 개발하였다. 이들은 또 공정 모델에 예측 모델식을 추가함으로써 필요한 매개 변수 추정치가 가능함을 보였다.

MacDonald와 Howat[10]은 단일 플래시 공정에 데이터 보정 기법을 적용하여, 그 적용성, 매개 변수의 추정 능력, 데이터 보정 효과 등을 검토한 바 있다. 이후 Serth 등[13]은 MacDonald와 Howat의 플래시 모델 및 알고리즘을 개선하고, 이를 gross error의 감지로 확장하였다. 이들은 모두 선형 근사를 이용하여 비선형 모델에 대한 알고리즘을 제시하였으며, 이러한 선형 근사의 효율성 및 적용성을 제시하였다.

실제 공정에 데이터 보정 기법을 적용한 예로는 Islam 등[14]이 열 분해 반응기에 대한 응용 결과를 보고한 바 있고, Pierucci 등[7]과 Canfield 등[5]이 대규모 석유 화학 공장에 데이터 보정과 최적화를 동시에 적용하여 좋은 결과를 얻었음을 보고하였다. 이러한 연구들은 화학공정의 상세 모델을 데이터 보정에 이용하여 얻을 수 있는 여러 가지 장점들을 실제로 보이고 있다.

2. 제안된 알고리즘

2-1. 데이터 보정

제안된 알고리즘은 Stephenson과 Shewchuck[12]이 제안한 방정식 중심 기법을 기본으로 한다. 데이터 보정 작업은 최적화 문제와 유사한 성격을 갖게 되므로, 상용 모사기에서 주로 채택되는 순차적 모듈 방식으로 처리하는 데는 많은 계산 시간과 반복계산이 필요하다. 그러나, 방정식 중심법에서는 공정의 모사와 최적화 작업이 동시에 이루어지므로 계산 효율면에서 매우 유리하고, 프로그램의 구현에 있어서도 그 확장성이 용이하다. 데이터 보정 문제는 다음과 같이 정의된다 [3, 11].

$$\min Q(\mathbf{x}) = \sum_{i=1}^m \frac{w_i (x_i^+ - x_i^-)^2}{V_{x_i}^2} \quad (1)$$

subject to $\mathbf{f}(\mathbf{z}) = \mathbf{f}(\mathbf{x}, \mathbf{y}) = 0$

변수에 대한 가중치 w_i 를 도입하여, 각각의 변수에 대한 공정 데이터의 신뢰성 경험치를 반영할 수 있도록 하였으며, 서로 다른 변수간의 공분산은 무시하였다. 이는 화학공정의 측정값에 대하여 일반적으로 적용되는 가정이며, 만일 공분산이 존재하는 경우 이를 확장하는 것은 어려운 일이 아니다. 제약 조건 $\mathbf{f}(\mathbf{z})$ 는 화학공정을 표현하는 모델 방정식들의 집합이며, 일반적인 물질 및 열 수치, 평형 관계식, 열역학적 상관식 등으로 표현된다.

식 (1)에 대해 Lagrangian을 구성하고

$$L(\mathbf{z}, \lambda) = Q(\mathbf{x}) + \lambda^T \mathbf{f}(\mathbf{z}) \quad (2)$$

이로부터 정지점 조건을 구해 보면 다음과 같은 식을 얻을 수 있다.

$$\begin{aligned} \frac{\partial L}{\partial \mathbf{x}} &= \frac{\partial Q}{\partial \mathbf{x}} + \left[\frac{\partial \mathbf{f}(\mathbf{x}, \mathbf{y})}{\partial \mathbf{x}} \right]^T \lambda = 0 \\ \frac{\partial L}{\partial \mathbf{y}} &= \left[\frac{\partial \mathbf{f}(\mathbf{x}, \mathbf{y})}{\partial \mathbf{y}} \right]^T \lambda = 0 \\ \frac{\partial L}{\partial \lambda} &= \mathbf{f}(\mathbf{x}, \mathbf{y}) = 0 \end{aligned} \quad (3)$$

식 (3)은 비선형 제약 조건인 세 번째 식 때문에 $(l+m+n)$ 차원의 비선형 방정식계가 된다. 이를 풀기 위해서는 비선형 풀이 기법을 이용할 수도 있으나, Hessian의 계산 부담을 없애기 위해 제안된 알고리즘에서는 제약 조건을 부분적으로 선형 근사하여 반복적으로 계산하도록 하였다. 즉, 제약 조건을 (x_i, y_i) 부근에서 Taylor 급수에 의한 선형 근사식을 전개하고, 각 항들을 정리하면 다음 식을 얻을 수 있다. 여기서 아래첨자 i 는 반복계산에서의 i 번째 구한 추정치를 의미한다.

$$\begin{aligned} \frac{\partial L}{\partial \mathbf{x}} &= \mathbf{A} \Delta \mathbf{x} + \left[\frac{\partial \mathbf{f}(\mathbf{x}, \mathbf{y})}{\partial \mathbf{x}} \right]^T \lambda - \mathbf{p} = 0 \\ \frac{\partial L}{\partial \mathbf{y}} &= \left[\frac{\partial \mathbf{f}(\mathbf{x}, \mathbf{y})}{\partial \mathbf{y}} \right]^T \lambda = 0 \\ \frac{\partial L}{\partial \lambda} &\approx \mathbf{f}(\mathbf{x}_i, \mathbf{y}_i) + \frac{\partial \mathbf{f}(\mathbf{x}_i, \mathbf{y}_i)}{\partial \mathbf{x}} \Delta \mathbf{x} + \frac{\partial \mathbf{f}(\mathbf{x}_i, \mathbf{y}_i)}{\partial \mathbf{y}} \Delta \mathbf{y} = 0 \end{aligned} \quad (4)$$

$$\text{where } A_{ki} = \begin{cases} 2w_k/V_{x_k}^2 & : \text{if } k=l \\ 0 & : \text{otherwise} \end{cases}$$

$$p_k = 2w_k(x_k^+ - x_{k,i})/V_{x_k}^2$$

이 식을 $\mathbf{z}=[\mathbf{x}|\mathbf{y}]$ 를 이용하여 행렬 형태로 다시 정리하면 다음과 같은 식을 얻는다.

$$\begin{bmatrix} \mathbf{J}(\mathbf{z}_i) & 0 \\ \mathbf{G} & \mathbf{J}^T(\mathbf{z}_i) \end{bmatrix} \begin{bmatrix} \Delta \mathbf{z} \\ \lambda \end{bmatrix} = \begin{bmatrix} -\mathbf{f}(\mathbf{z}_i) \\ \mathbf{h} \end{bmatrix} \quad (5)$$

$$\text{where } G_{kk} = \begin{cases} 2w_k/V_{x_k}^2 & : \text{if } k \text{ is measured variable} \\ 0 & : \text{otherwise} \end{cases}$$

$$h_k = \begin{cases} 2w_k(z_k^+ - z_{k,i})/V_{x_k}^2 & : \text{if } k \text{ is measured variable} \\ 0 & : \text{otherwise} \end{cases}$$

$$\mathbf{J}(\mathbf{z}_i) = \left[\frac{\partial \mathbf{f}}{\partial \mathbf{z}} \right]_{\mathbf{z}_i} = \left[\frac{\partial \mathbf{f}}{\partial \mathbf{x}} \frac{\partial \mathbf{f}}{\partial \mathbf{y}} \right]_{\mathbf{z}_i}$$

위 식의 수렴 조건은 다음과 같이 설정하였다.

$$\begin{aligned} \sum_{k=1}^{m+n} \left| \frac{\Delta z_{i,k}}{z_{i,k}} \right| &< \text{TOLERANCE}_{\text{relative}} \\ \mathbf{f}(\hat{\mathbf{z}}_i)^T \mathbf{f}(\hat{\mathbf{z}}_i) &< \text{TOLERANCE}_{\text{residual}} \end{aligned} \quad (6)$$

우선 주어진 입력으로부터 Jacobian 계산을 통해 각 단위 장치간의 연결 관계를 결정하고, 식 (5)로 표현되는 기준 방정식을 구성한다. 이때, Jacobian의 희소성을 파악하여 이후의 반복 계산에서는 이의 구조를 고정시킨다. 매 단계에서 Jacobian의 계산은 수치 미분을 이용하였다. 수렴 조건이 만족될 때까지 계산을 반복하게 되며, 수렴해를 얻은 후에는 gross error 감지를 수행한다. 만일 gross error가 존재하면 이 변수를 미측정 변수로 소거하고, 각 변수들의 관찰 가능성을 확인하여 전체 계산을 반복하도록 구성되어 있다. 최종적으로 gross error가 없는 해를 얻으면, 각 측정 변수 및 그 보정치, 미측정 변수들의 분산을 계산하는 후처리 단계를 거쳐 결과를 출력한다. 다음의 Fig. 1에 제안된 데이터 보정 절차를 개념적으로 나타내었다.

본 연구에서는 선형 근사를 통해 구성된 기준 방정식들에 대해 QP

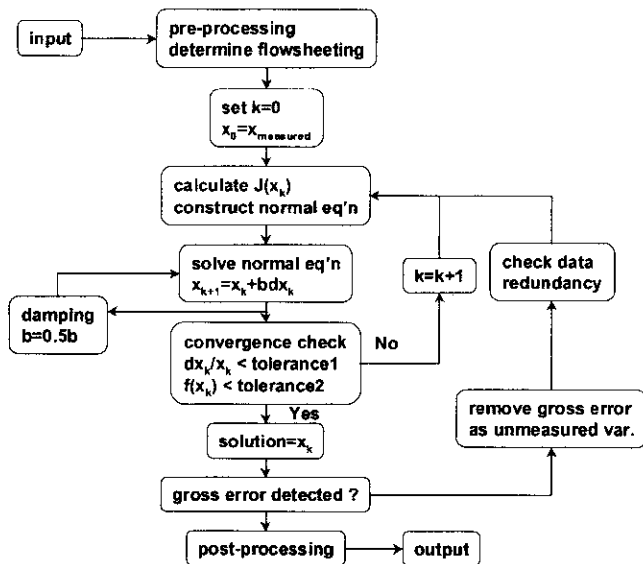


Fig. 1. Algorithm for data reconciliation.

법에서 일반적으로 채택되는 풀이 절차나 동시 방법 대신에 직접 Newton 방법을 적용하였다. 이를 통해 얻을 수 있는 장점은 불필요한 행렬간의 곱셈 계산 및 역행렬 계산을 피할 수 있다는 점이다.

2-2. Gross Error 감지-교차점 테스트(Point Test)

측정치와 데이터 보정을 통해 얻은 추정치가 정규 분포를 따른다면 다음과 같은 확률 밀도를 갖는다.

$$f(x_i^+) = \frac{1}{\sigma_{x_i^+} \sqrt{2\pi}} \exp \left[-\frac{1}{2} \frac{(x_i - x_i^+)^2}{\sigma_{x_i^+}^2} \right] \quad (7)$$

$$f(\hat{x}_i) = \frac{1}{\sigma_{\hat{x}_i} \sqrt{2\pi}} \exp \left[-\frac{1}{2} \frac{(x_i - \hat{x}_i)^2}{\sigma_{\hat{x}_i}^2} \right] \quad (8)$$

식 (7)과 (8)로 표현되는 각각의 확률 밀도의 예를 Fig. 2와 같은 그림으로 나타내 볼 수 있다. 이 그림은 측정치의 확률 밀도의 평균을 추정치로 이동시켜 함께 표현한 것이다. 이때 두 분포 사이에 두 개의 교차점이 존재하게 되는데 이 교차점 사이의 영역은 무작위 오차의 영역으로, 그 밖의 영역은 gross error의 존재에 의한 영역으로 생각할 수 있다[15]. 즉, 측정값이 이 두 교차점의 사이 구간에 존재하는가의 여부에 따라 이 변수가 gross error인지를 판단하는 기준으로 삼

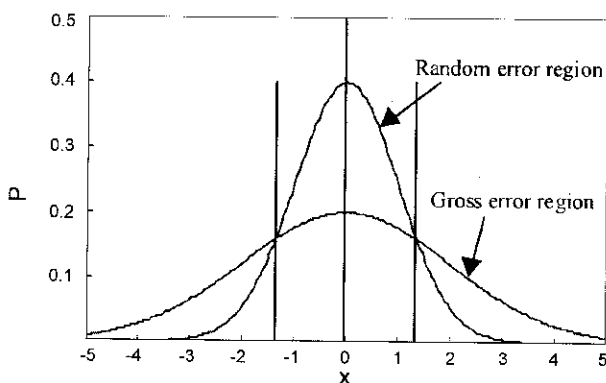


Fig. 2. Probability distribution of measured & estimated value with same mean.

Table 1. Criterion values for point test

γ_i	z_{pci}
0.05	2.450
0.1	2.156
0.2	1.831
0.3	1.626
0.4	1.477
0.5	1.359
0.7	1.146
1.0	1.00

는다. 위의 식 (7)과 (8)에서 측정치와 추정치로 치환하여 두 식을 같다고 놓고서 추정치와 측정치의 표준 편차의 비를 γ_i , 추정 오차를 ε_i 으로 표현하고 정리하면 다음의 식을 얻을 수 있다.

$$|\varepsilon_i|^2 = 2\sigma_{x_i}^2 \frac{\ln \gamma_i}{\gamma_i^2 - 1} \quad (9)$$

where, $\gamma_i = \sigma_{x_i} / \sigma_{x_i^+}$

$$|\varepsilon_i| = |x_i - \hat{x}_i|$$

앞의 두 식으로부터 다음과 같은 판별 기준을 설정할 수 있다.

$$\text{if } z_{p_i} \left(= \frac{|\varepsilon_i|}{\sigma_{x_i}} \right) \leq z_{pci} \left(= \sqrt{\frac{2 \ln \gamma_i}{\gamma_i^2 - 1}} \right), \quad \text{then variable } i \text{ is not a gross error} \quad (10)$$

if $z_{p_i} > z_{pci}$, then variable i is a gross error

Table 1에서 보는 바와 같이 이 교차점 테스트는 γ_i 의 값에 따라 서로 다른 기준값이 적용된다. 이 방법은 각 변수에 대한 보정 전후의 표준 편차비가 1에 가까울수록 작은 기준값이 적용된다. 따라서, 고정된 한 값을 기준으로 삼는 기존의 측정치 테스트에 비해 보정 편차에 따른 상대적 가중을 두므로, gross error 감지의 선별도를 높일 수 있다. 이 교차점 테스트는 기존의 방법들과 결합되어, gross error 존재의 재확인 및 기존의 방법으로는 감지되지 않은 gross error의 존재를 제시하는데 이용할 수 있다.

그러나, 수치 실험 결과 대규모 방정식계에서는 위의 기준값들이 너무 작아 지나치게 많은 gross error가 감지되는 결과를 보였다. 따라서, 공정 변수의 수에 따라 위의 기준 조건을 적절히 완화시킬 수 있는 매개 변수를 도입하였다.

식 (10)을 다음과 같이 다시 표현해 보자.

$$\text{if } z_{p_i} \left(= \frac{|\varepsilon_i|}{\sigma_{x_i}} \right) \leq z_{pci} \left(= k_{p_i} u_{1-\alpha/2} \sqrt{\frac{2 \ln \gamma_i}{\gamma_i^2 - 1}} \right),$$

then variable i is not a gross error

if $z_{p_i} > z_{pci}$,

then variable i is a gross error (11)

이 식에서 $u_{1-\alpha/2}$ 는 측정치 테스트에서와 마찬가지로 주어진 유의 수준 α 에 대한 정규 분포의 해당 좌표이다. 이를 사용한 이유는 $\gamma_i = 1$ 인 경우의 교차점 테스트의 기준값을 측정치 테스트와 같게 설정하기 위해서이다. k_{p_i} 는 교차점 테스트에 대해 임의로 설정한 가중치이며, 이에 따라 감지 성능이 크게 좌우될 수 있다. 최적의 k_{p_i} 값은 변수의 수, 각 변수의 측정치 분산 등 대상 시스템의 특성에 따라 달라진다. 초기 k_{p_i} 는 다음과 같은 방법으로 설정할 수 있다. 우선,

$$\sigma_{a_i}^2 = \sigma_{x_i^+}^2 - \sigma_{\hat{x}_i}^2 = \left(\frac{1 - \gamma_i^2}{\gamma_i^2} \right) \sigma_{x_i}^2 \quad (12)$$

또한 다음의 관계가 성립한다.

$$\frac{\varepsilon_i}{\sigma_{\varepsilon_i}} = \sqrt{\frac{\gamma_i^2}{1-\gamma_i^2}} \frac{\varepsilon_i}{\sigma_{x_i}} \sim N(0, 1) \quad (13)$$

위의 관계로부터 k_{p_i} 는 다음과 같이 구해진다.

$$k_{p_i} = \sqrt{\frac{1-\gamma_i^2}{\gamma_i^2}} \quad (14)$$

이 교차점 테스트는 단순히 식 (14)를 이용하여도 측정치 테스트보다 좋은 감지 성능을 보이지만, 대상 공정의 특성에 따라 식 (14)를 k_{p_i} 의 초기값으로 하여 실제 공정에 대한 gross error 감지 성능이 향상 되도록 조정해 나갈 수 있다. 또한, 변수들간에 공선형성을 보여 이들 변수 중 하나에 gross error가 발생하는 경우 측정치 테스트의 경우에는 이를 구분할 수 없으나 이 교차점 테스트에는 공선형의 문제가 발생하지 않는 장점이 있다.

3. 시스템 구현

화학공정의 정상상태 모델을 이용한 데이터 보정을 수행하기 위해 전체 프로그램의 구조는 Fig. 3과 같이 구현하였다. 구현된 시스템은 FORTRAN으로 작성되었으며, 화학공정의 단위 장치들에 대한 정상상태 모델 및 열역학적 물성치 계산을 위한 데이터 베이스를 갖추고 있다.

화학공정의 특성 중 하나는 그 Jacobian 행렬의 희소성이 매우 크다는 점이다. 행렬의 희소성을 이용하여 필요한 메모리의 양을 줄이도록 고안되었으며, 계산의 효율성을 높이기 위해 희소 행렬 알고리즘을 적용하였다. 구현된 시스템의 또 한 가지 특징은 공정 모사가 가능하다는 점이다. 여기서 공정 모사를 주어진 설계 변수의 규정값(specification value)을 만족하도록 하는 공정 변수들의 값을 계산하는 작업이다. 구현된 시스템에서 각 단위 장치에 대한 자유도를 만족하도록 설계 변수(구현된 시스템 내에서는 측정 변수로 간주된다)를 규정하고, 계산하고자 하는 변수들을 미측정 변수로 설정하면 이러한 모사를 수행할 수 있다.

3-1. 수치 실험 방법

데이터 보정 성능을 분석하기 위해 위의 예제에 대해 Monte Carlo 모사를 수행하였다. 본 연구에서는 미리 알고 있다고 가정한 참값으로부터 측정값을 생성하기 위해 난수를 이용하였다. 즉, 각각의 예제에 주어진 변수값을 참값으로 가정하고 여기에 인위적으로 생성시킨 무작위 오차를 더하여 측정값을 생성시킨다. 이 무작위 오차는 각 변

수의 측정치 표준 편차와 같은 표준 편차를 갖고 평균은 0인 Gaussian Noise이다. 또한 균일 분포를 이용해 gross error의 크기 및 변수 위치를 결정하였다. 이렇게 생성된 측정치는 다음과 같이 표현된다.

$$x^+ = \hat{x} + e_R + e_G$$

where, $e_R \sim N(0, s_x^2)$: random error
 $e_G \sim \text{uniform distribution}$: gross error

여기서 gross error를 나타내는 e_G 항의 크기는 참값의 $\pm 10\% \sim \pm 100\%$ 가 되게 하였다. 즉,

$$0.1|\hat{x}| \leq |e_G| \leq 1.0|\hat{x}| \quad (15)$$

4. 사례연구

4-1. 비평형 플래시 공정의 데이터 보정

비평형 단일 플래시 증류 공정에 대하여 본 연구에서 제안한 데이터 보정 알고리즘을 적용하여 보았다. Table 2는 MacDonald와 Howat이 사용하였던 예제의 규정값들이며, Table 3은 주어진 입력 조건에 따른 비평형 플래시 계산 결과이다. 이 계산은 Table 2에 주어진 변수들의 값을 측정 변수로 규정하고, 그 밖의 변수들을 미측정 변수로 규정하여 제안된 시스템 내에서 계산되었다. 이 결과는 데이터 보정 효과 없이 설계 계산, 즉 공정 모사를 수행한 것이다.

각 측정 변수의 표준 편차는 온도는 2°R , 압력과 유량은 측정치의 5%, 에너지 유량 및 조성은 측정치의 10%로 가정하여 앞에서 설명한 Monte Carlo Simulation을 100번 수행하였고, 그 결과를 MacDonald와 Howat의 결과와 비교하여 Table 4에 나타내었다. 여기서 gross error는 없는 것으로 간주하였다.

이 결과에서 주목할 사항은 온도, 압력, 열 유량 등의 변수에 대한 데이터 보정 효과가 다른 변수들(유량, 조성)에 비해 상대적으로 작다는 사실이다. 온도, 압력 등의 변수들은 모델 방정식의 열역학 물성치를 계산하는데 큰 영향을 미치므로 각각의 변수에 포함되어 있는 오차가 상호간에 평형 관계식, 열 수지식 등으로 연결되어 서로 강한 영향을 주기 때문이다. 이 변수들에 대한 신뢰도의 향상 효과(표

Table 2. Non-equilibrium flash condition

Flash feed condition	
Flow rate(lbmol/hr)	1,000.0000
Temperature(R)	540.3000
Pressure(psia)	85.6500
Mole fraction	Propane 0.6000
	n-Butane 0.2000
	n-Pentane 0.2000
Flash pressure(psia)	21.7600
Heat input(mmBtu/hr)	-2.0332
Vaporization efficiency	0.8500

Table 3. Result of non-equilibrium flash calculation

Product condition	
Liquid flow rate(lbmol/hr)	788.6831
Flash temperature(R)	468.1000
Mole fraction	Propane 0.5205
	n-Butane 0.2311
	n-Pentane 0.2484
Vapor flow rate	211.3169
Mole fraction	Propane 0.8965
	n-Butane 0.0840
	n-Pentane 0.0195

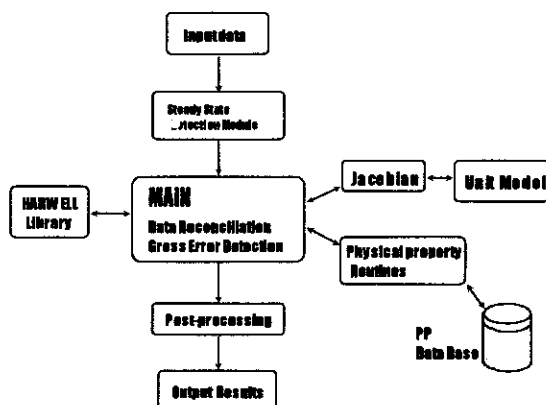


Fig. 3. System structure.

Table 4. Comparison of the result with MacDonald and Howat's

		Proposed Algorithm				MacDonald & Howat	
		NGOOD	DEV	DEVA	STDDEVA	NGOOD	STDDEVA
1	F	66	43.785	23.907	31.428	68	30.000
2	T_F	55	1.477	1.290	1.712	62	1.8
3	P_F	64	3.685	2.739	2.846	-	-
4	z_1	90	0.054	0.009	0.013	86	0.014
5	z_2	72	0.015	0.008	0.010	74	0.011
6	z_3	73	0.016	0.007	0.010	67	0.012
7	TFL	57	1.704	1.616	1.978	44	1.8
8	PFL	51	0.928	0.924	1.080	60	1.088
9	Q	49	0.237	0.242	0.285	68	0.291
10	θ	-	-	0.042	0.055	-	0.037
11	L	63	33.113	23.866	30.909	63	30.759
12	x_1	86	0.040	0.010	0.014	73	0.023
13	x_2	78	0.020	0.009	0.012	67	0.012
14	x_3	76	0.021	0.009	0.012	63	0.017
15	V	50	7.752	7.986	10.343	67	8.664
16	y_1	96	0.071	0.004	0.006	83	0.014
17	y_2	64	0.006	0.004	0.005	44	0.006
18	y_3	74	0.002	0.001	0.001	48	0.002
Mean eff.			0.8477			0.85	
Minimum			0.7307			0.71	
Maximum			0.9889			0.99	
Reduction of Q			54.85%				

STDDEVA=Standard Deviation of Estimates. NGOOD(%)=Number of Estimations Better than Measured Values.

DEV=Mean of Absolute Deviations of Measurements from True Values. DEVA=Mean of Absolute Deviations of Estimates from True Values.

준 편차의 감소)도 다른 변수들에 비해 작은 편이다.

대체적으로 유량의 보정 효과는 두 방법이 비슷하고, 온도 및 압력은 MacDonald와 Howat의 방법이, 조성은 본 연구에서 제안된 방법의 보정 효과가 좋게 나타났다. 매개 변수 추정에 있어서는 평균은 같고, 그 추정 범위는 제안된 알고리즘이, 표준 편차는 MacDonald와 Howat의 결과가 좋다.

이 결과에 대한 평가에 앞서 주목할 점은 각 방법에서 사용한 모델이 얼마나 실제 공정을 잘 표현하는지에 대한 것이다. 본 연구에서는 증발효율(vaporization efficiency)을 이용하여 모델을 구성하였고, MacDonald와 Howat은 Murphree efficiency를 이용하였다. Murphree efficiency를 사용하는 경우는 (C-1)개의 평형 관계식만이 독립적이므로 데이터 보정에 이용할 수 있는 하나의 관계식을 잃게 되고, 전체 시스템의 자유도가 하나 더 증가한다. 이 문제를 해결하기 위해 MacDonald와 Howat은 비평형 플래시 온도가 평형 플래시 온도와 같다는 가정을 하나 더 도입하였으며, 그 결과로 입력 스트림의 압력에 대한 보정이 없이 값이 고정된다. 이 때문에 이들의 결과가 온도, 압력 등에 대해 본 연구에서 제안한 방법보다 좋은 결과를 나타낸다. 그러나, 추가로 설정된 이러한 가정은 상당히 임의적이며, 실제 공정의 상태를 잘 표현한다고 보기는 어렵다.

이상에서 보는 바와 같이 본 연구에서 이용한 증발 효율 모델은 이용 가능한 모델식을 최대한 활용하며, 불필요한 가정의 도입이 필요 없으므로 실제 공정의 물리적 의미를 더욱 잘 표현한다. 또한 평형 관계식의 Jacobian을 계산하는데 있어서도 증발효율을 이용하는 것이 훨씬 효율적이다.

4.2. 교차점 테스트(Point Test)

앞의 단일 플래시 공정을 대상으로 측정치 테스트와 교차점 테스트의 gross error 감지 성능을 비교해 보았다. 각각의 경우에 대해 100 번의 Monte Carlo simulation을 통하여 gross error 감지 실험을 수행하고, 이 중 플래시 온도 T에 +92%의 gross error가 존재하는 경우

Table 5. Measurement test for a non-equilibrium flash problem

GR ERR. VARIABLE=9		
ADDED GR SIZE=+92.42%		
Measurement test		
Z_i Values=		
2 T_F	4.9155	Gross Error Detected
6 Z_3	-3.5099	Gross Error Detected
9 Q	4.8661	Gross Error Detected
Z_i Criterion Value=2.9594		
LARGEST Z_i VAR. SUSPECTED=2		$Z_i=4.9155$

에 대한 측정치 테스트와 교차점 테스트의 결과를 Table 5와 6에 나타내었다. 먼저 Table 5의 측정치 테스트 결과를 보면 gross error로 감지된 변수는 3개(T_F , z_3 , Q)이고, 이 중 T_F 와 Q는 에너지 수지식 하나에만 나타나며, 서로 공선형의 관계를 갖고 있다. 그러나, 선형 근사의 오차 때문에 약간의 차이가 발생하며, 최대의 z_{ai} 값을 보이는 변수는 T_F 이다. 이 변수를 미측정 변수로 간주하고 다시 데이터 보정 및 측정치 테스트를 수행하면 더 이상의 gross error가 감지되지 않는다. 하지만 잘못된 gross error의 감지로 인해 Q에 존재하는 오차는 그대로 남게 되며, 이에 의한 오차가 다른 변수들에 전달되어 데이터 보정 후의 총 오차합이 오히려 증가하게 된다.

Table 6은 본 연구에서 제안된 교차점 테스트에 의한 gross error 감지 실험 결과이다. 교차점 테스트에 의해 감지된 gross error는 T_F 와 Q 두 개로서, 측정치 테스트의 경우보다 오차 원인 후보가 줄었다. 또한, 측정치 테스트와는 달리 근소한 차이나마 최대의 z_{pi}/z_{pci} 값을 보이는 변수는 Q이므로 이 변수를 최종 gross error로 간주, 제거하게 된다. 그 결과 데이터 보정은 측정치 테스트의 경우보다 우수하다. 모든 경우에 대해 교차점 테스트가 이러한 공선형성을 해결할 수 있는 것은 아니지만, 측정치 테스트보다는 선별도가 높음을 알 수 있다. 이 예제의 경우에는 식 (14)를 이용하여 k_{pi} 의 초기값을 정하였으

Table 6. Point test for the non-equilibrium flash problem

GR ERR. VARIABLE=9					
ADDED GR SIZE=+92.42%					
Point test					
	Z_{pi}	Z_{pci}	Z_{pi}/Z_{pci}		STDDEV/STDDEVA
2 T_F	3.1838	2.0869	1.5256	Gross Error Detected	0.8393
9 Q	1.9751	1.2472	1.5836	Gross Error Detected	0.9266
LARGEST Z_{pi} VAR. SUSPECTED=9			$Z_{pi}/Z_{pci}=1.5836$		

STDDEV=Standard Deviation of Measured Values

Table 7. Performance of measurement test for the flash problem

	Measurement test $\alpha=0.05$			
Accuracy of measurements	$0.25 \times$ STDDEV	$0.5 \times$ STDDEV	$1.0 \times$ STDDEV	$1.2 \times$ STDDEV
Gross errors identified	72	68	48	39
Gross errors not detected	28	32	52	61
Wrong identification	33	23	28	17
Selectivity*	0.686	0.747	0.640	0.696
Efficiency estimation				
Mean	0.8232	0.8438	0.8374	0.8551
Minimum	0.0180	0.0919	0.2288	0.1731
Maximum	1.5741	1.3897	1.4433	1.7420
Std. dev. of efficiency	0.030	0.040	0.057	0.061

*Selectivity=No. of Gross Errors Identified/(No. of Gross Errors Identified+No. of Wrong Identification)

Table 8. Performance of point test for the flash problem

	Point test $\alpha=0.05$			
Accuracy of measurements	$0.25 \times$ STDDEV	$0.5 \times$ STDDEV	$1.0 \times$ STDDEV	$1.2 \times$ STDDEV
Gross errors identified	82	70	46	32
Gross errors not detected	18	30	54	68
Wrong identification	16	15	29	14
Selectivity	0.837	8.24	0.613	0.696
Efficiency estimation				
Mean	0.8363	0.8542	0.8914	0.8750
Minimum	0.1068	0.7661	0.0538	0.5678
Maximum	1.4655	1.3865	2.6488	1.6082
Std. dev. of efficiency	0.027	0.039	0.062	0.068

나 실제 응용에 있어서는 여러 가지 상황을 살펴 gross error가 발생할 가능성이 높은 측정변수에 대해 도입된 파라미터인 k_{pi} 의 값을 조정하여 가중치를 변화시킬 수 있다. 이 경우에는 T_F 가 Q 대신 선택되는 것을 원하지 않으므로 Q에 대한 파라미터 k_{pi} 의 값을 줄임으로써 z_{pi}/z_{pci} 값을 증가시켜 선별도를 높일 수 있다.

Table 7과 8은 100번의 gross error 감지 실험 실행 결과에 대한 감지 성능을 나타낸다. 두 방법 모두 유의 수준 $\alpha=0.05$ 로 설정하였고 감지된 gross error의 회수, 감지하지 못한 gross error 회수, 그리고 실제로는 gross error가 아니지만 gross error로 잘못 감지된 경우의 회수를 나타내었다. 이때 대상 시스템 측정값의 정확도에 따른 영향을 알아보기 위해, 측정치의 분산을 변화시켜 보았다. 두 방법 모두 측정치의 정확도가 떨어질수록 감지 성능이 나빠지는 경향을 보였다. 측정값의 정확성이 높을 때는 교차점 테스트가 우수한 성능을 보였으나, 반대로 측정값의 정확도가 낮을 때는 측정치 테스트가 우수한 결과를 보인다. 이를 고려하면 gross error 감지를 위한 교차점 테스트 혹은 측정치 테스트를 이용할 때는 대상 시스템의 정확도에 따라 선택의 여지가 있다. 그러나, 이 문제의 경우에 기준 표준 편차는 온도 2R, 압력과 유량은 5%, 에너지량 및 조성은 10%를 이용하였다.

이러한 표준 편차는 실제 공정의 일반적인 측정치보다 상당히 큰 값이다. 따라서, 실제 공정에서의 응용성은 교차점 테스트가 더 유리한 것으로 판단된다.

5. 결 론

비선형 방정식계로 표현되는 화학공정 모델을 이용한 데이터 보정 및 gross error 감지 기법을 제안하였고 이를 비평형 단일 플래시 증류 공정에 적용하여 그 효용성을 입증하였다.

본 연구에서 제안된 데이터 보정 방법은 선형 근사를 통해 구성된 기준 방정식들에 대해 직접 Newton 방법을 적용하여 불필요한 행렬 간의 곱셈 계산 및 역행렬 계산을 피할 수 있다는 장점이 있으며, 제안한 교차점 테스트는 기존의 측정치 테스트가 갖는 공선형성의 문제를 부분적으로 해결하였으며, 단일 감지 기법으로서 측정치 테스트보다 우수한 감지 성능을 보였다.

그리고 제안된 데이터 보정 및 gross error 감지 기법을 통합하여 하나의 시스템으로 구축하였다. 이 시스템은 기본적인 단위 공정의 모델, 물성치 루틴 등을 갖추고 있으므로, 하나의 독립된 시스템으로 사용이 가능하다. 또한 구현된 시스템은 공정 모델의 추가 및 수정이 자유로우므로 따라서 대규모 화학공정에도 효율적인 적용이 가능할 것으로 기대된다.

감 사

본 연구는 과학기술정책관리연구소(STEPI)의 엔지니어링 핵심공동 기반기술 개발과제의 지원금에 의한 것입니다.

사용기호

- l : number of constraint equations
- m : number of measurable variables
- n : number of unmeasured variables
- V_x : covariance matrix of measurements
- w : weight vector
- x =vector of measurable values ($m \times 1$)
- y =vector of unmeasured value ($n \times 1$)
- $z=[x|y]$
- z_a : criterion value for measurement test
- z_p : criterion value for point test

그리스 문자

- α : level of significance
- ϵ : measurement error
- γ : the ratio of standard deviation of estimated value and measured value
- λ : Lagrangian multiplier vector
- σ : standard deviation

상첨자

- $+$: measurement value
- \wedge : estimated or corrected value
- T : matrix transpose

하첨자

- i : index in vector or matrix

참고문헌

1. Hlavacek, H.: *Comput. Chem. Engng.*, **1**, 75(1977).
2. Tamhane, A. C. and Mah, R. S. H.: *Technometrics*, **27**(4), 409(1985).
3. Crowe, C. M., Campos, Y. A. G. and Hrymak, A.: *AIChE J.*, **29**(6), 881(1983).
4. Crowe, C. M.: *AIChE J.*, **34**(4), 541(1988).
5. Yang, S. F. and Canfield, F. B.: Proceedings of PSE'91, **II**, 22(1991).
6. Nair, P. and Canfield, F. B.: SICHEM'92, **3**, Seoul, Korea(1992).
7. Bussani, G., Chiari, M., Grottoli, M. G., Pierucci, S., Faravelli, T. and Ricci, G.: Proceedings of ESCAPE-5, **19**, 299(1995).
8. Britt, H. I. and Luecke, R. H.: *Technometrics*, **15**(2), 233(1973).
9. Knepper, J. C. and Gorman, J. W.: *AIChE J.*, **26**(2), 260(1980).
10. MacDonald, R. J. and Howat, C. S.: *AIChE J.*, **34**(1), 1(1988).
11. Veverka, V.: *Appl. Math. Comput.*, **49**, 141(1992).
12. Stephenson, G. R. and Shewchuk, C. F.: *AIChE J.*, **32**(2), 247(1986).
13. Serth, R. W., Srikanth, B. and Maronga, S. J.: *AIChE J.*, **39**(10), 1726(1993).
14. Islam, K. A., Weiss, G. H. and Romagnoli, J. A.: Proceeding of ESCAPE-4, **18**, 217(1994).
15. Tjoa, I. B. and Biegler, L. T.: *Comput. Chem. Engng.*, **15**(10), 679(1991).