UKACC International Conference on Control 2012
Cardiff, UK, 3-5 September 2012

Branch and Bound Method for Globally Optimal Controlled Variable Selection

Vinay Kariwala
ABB Global Industries & Services Ltd
Mahadevpura, Bangalore 560048, India
Email: vinay.kariwala@in.abb.com

Lingjian Ye
Ningbo Institute of Technology
Zhejiang University
315100, Ningbo, Zhejiang, China
Email: lingjian.ye@gmail.com

Yi Cao
School of Engineering
Cranfield University
Cranfield, Bedford MK43 0AL, UK
Email: Y.Cao@cranfield.ac.uk

***Abstract*—For selection of controlled variables (CVs) in self-optimizing control, various criteria have been proposed in the literature. These criteria are derived based on local linearization of the process model and the necessary conditions of optimality (NCO) at a nominally optimal operating point. Recently, a novel CV selection framework has been proposed by Ye et al. [1] by converting the CV selection problem into a regression problem to approximate the NCO globally over the entire operation region. In this approach, linear combinations of a subset of available measurements are used as CVs. The subset selection problem is combinatorial in nature redering the application of the globally optimal CV selection method to large-scale processes difficult. In this work, an efficient branch and bound (BAB) algorithm is developed to handle the computational complexity associated with the selection of globally optimal CVs. The proposed BAB algorithm identifies the best measurement subset such that the regression error in approximating NCO is minimized. This algorithm is applicable to the general regression problem. The efficiency and effectiveness of the proposed BAB algorithm is demonstrated through a binary disdillation column case study.**

I. Introduction

The selection of controlled variables (CVs) from available measurements is an important task during the design of control systems. For CV selection, several methods have been proposed in the literature. Skogestad [2] introduced the concept of self-optimizing control for selection of CVs based on process economics. In this approach, CVs are selected such that in presence of disturbances, the loss incurred in implementing the operational policy by holding the selected CVs at constant setpoints is minimal, as compared to the use of an online optimizer. The advantages of self-optimizing control approach for CV selection has been demonstrated through several case studies; see *e.g.* [3] for an overview.

The choice of CVs based on the general non-linear formulation of self-optimizing control requires solving large-dimensional non-convex optimization problems [2]. To quickly pre-screen alternatives, exact local methods with worst-case [4] and average loss minimization [5] have been proposed. These local methods are useful for selecting a subset or linear combinations of available measurements as CVs, where the latter approach provides lower losses. Recently, explicit solutions to the problem of finding locally optimal measurement combinations have been proposed [6], [5], [7]. Hu *et al.* [8] have proposed a local method to explicitly handle the input and output constraints during CV selection.

The available CV selection criteria are derived based on local linearization of the process model. Recently, a globally optimal CV selection framework has been proposed by Ye *et al.* [1], [9]. In this framework, the CV synthesis problem is converted into a regression problem using CVs as measurement combinations to approximate the Necessary Conditions of Optimality (NCO) globally over the entire operation region. It has been proven that the average loss is globally minimized when the regression error is minimal over the entire operation region and the measurement combinations as CVs are perfectly controlled at zero. A number of linear and nonlinear regression models have been proposed to approximate the NCO. Case studies showed that all these models are able to significantly reduce the loss, as compared to using CVs designed by using existing local methods.

In general, CV selection is a combinatorial problem. For selection of individual measurements as CVs, a number of efficient branch and bound (BAB) approaches, called bidirectional BAB (B^3), have been developed for various local criteria to address the combinatorial issue [18], [10], [11]. These BAB algorithms are not required for the selection of individual measurements as globally optimal CVs, as the selection is not combinatorial any more because approximations of individual gradients are not correlated and can be solved separately. However, to select measurement combinations, the combinatorial difficulty still exists for the global CV selection problem.

It is known that the use of combinations of a few measurements as CVs often provide similar loss as the case where combinations of all available measurements are used [6], [5], [7], [9]. Though the former approach results in control structures with lower complexity, it gives rise to a combinatorial optimization problem involving selection of measurements, whose combinations can be used as CVs. For local self-optimizing control methods, partially bidirectional BAB (PB^3) methods have been proposed to solve this combinatorial problem efficienctly [10], [11]. In this work, the framework is further extended to measurement subset selection for synthesis of globally optimal CVs chosen as linear combinations of measurements. It is proven that the selection criterion is equivalent to a quadratic problem, for which a standard BAB

978-1-4673-1558-6/12/$31.00 ©2012 IEEE

algorithm [12] exists. The standard algorithm is improved into a downwards BAB algorithm. The efficiency and effectiveness of the proposed BAB algorithm is demonstrated through a distillation case study [13].

II. Local Methods for Self-Optimizing Control

Consider that the steady-state economics of the plant is characterized by the scalar objective function $J(u,d)$, where $u \in \mathbb{R}^{n_u}$ and $d \in \mathbb{R}^{n_d}$ are inputs and disturbances, respectively. The optimal operation policy is to update u according to d, which usually requires the use of an online optimizer. For this case, let the optimal value of the objective function be denoted as $J_{\text{opt}}(d)$. A simpler strategy involves indirect adjustment of u using a feedback controller. In this case, the feedback controller manipulates u to hold the CVs c close to their specified setpoints. Here, in addition to d, J is also affected by the error e in implementing the constant setpoint policy, which results due to uncertainty and measurement noise. The suboptimal objective functional value under the second strategy is denoted as $J_c(e,d)$. Then, the worst-case and average losses due to the use of the suboptimal strategy are given as

$$\text{Worst-case loss} = \max_{e\in\mathcal{E}} \max_{d\in\mathcal{D}} (J_{\text{opt}}(d) - J_c(e,d)) \quad (1)$$

$$\text{Average loss} = E[J_{\text{opt}}(d) - J_c(e,d)] \quad (2)$$

$$(3)$$

where $\mathcal{D}$ and $\mathcal{E}$ represent the sets of allowable disturbances and implementation errors, respectively, and E is the expectation operator. Self-optimizing control is said to occur, when we can achieve an acceptable loss by holding the CVs close to their setpoints without the need to reoptimize when disturbances occur [2]. Based on this concept, the appropriate CVs can be selected by comparing the losses for different alternatives.

As mentioned earlier, the use of nonlinear formulation of self-optimizing control is difficult. Hence, some local methods were developed to estimate the losses defined in (1) and (2) by linearising the process model around the normally optimal operating point as follows:

$$y = G^y\, u + G_d^y\, W_d\, d + W_e\, e \quad (4)$$

where $y \in \mathbb{R}^{n_y}$ denotes the process measurements and $e \in \mathbb{R}^{n_y}$ denotes the implementation error, which results due to measurement and control errors. Here, the diagonal matrices W_d and W_e contain the expected magnitudes of disturbances and implementation error, respectively. The CVs $c \in \mathbb{R}^{n_u}$ are given as

$$c = H\, y = G\, u + G_d\, W_d\, d + H\, W_e\, e \quad (5)$$

where H is a selection or combination matrix and

$$G = H\, G^y, \quad G_d = H\, G_d^y \quad (6)$$

It is assumed that $G \in \mathbb{R}^{n_u \times n_u}$ is invertible. This assumption is necessary for integral control. When d and e are assumed to be uniformly distributed over the set

$$\left\| \begin{bmatrix} d^T & e^T \end{bmatrix}^T \right\|_2 \leq 1 \quad (7)$$

the local worst-case and average losses are given as [4], [5]:

$$L_{\text{worst}}(H) = 0.5\bar{\sigma}^2 \left(J_{uu}^{1/2} (H\, G^y)^{-1} H\, Y \right) \quad (8)$$

$$L_{\text{average}}(H) = \frac{1}{6(n_y + n_d)} \left\| J_{uu}^{1/2} (H\, G^y)^{-1} H\, Y \right\|_F^2 \quad (9)$$

where $\bar{\sigma}$ and $\|\cdot\|_F$ denote the maximum singular value and Frobenius norm, respectively, and

$$Y = \begin{bmatrix} (G^y\, J_{uu}^{-1}\, J_{ud} - G_d^y)\, W_d & W_e \end{bmatrix} \quad (10)$$

with $J_{uu} = \frac{\partial^2 J}{\partial u^2}$ and $J_{ud} = \frac{\partial^2 J}{\partial u \partial d}$, evaluated at the nominal operating point. In comparison with worst-case loss, the selection of CVs is preferred through minimization of average loss, as the worst-case may not occur frequently in practice [5].

When individual measurements are selected as CVs, H can be considered to be a selection matrix. Instead of using individual measurements, it is possible to use combinations of measurements as CVs. For this case, Alstad et al. [7] has recently proposed an explicit expression for H, which minimizes the L_{average} in (9) and is given as

$$H^T = (YY^T)^{-1} G^y ((G^y)^T (YY^T)^{-1} G^y)^{-1} J_{uu}^{1/2} \quad (11)$$

As shown by Kariwala et al. [5], the H in (11) also minimizes L_{worst} in (8). The locally optimal combinations of all the available measurements, which can be used as CVs can be found using (11).

III. Globally Optimal Method

The local methods [4], [5] are based on linearization around the nominally optimal operating point. Therefore, the identified CVs are only locally optimal. To derive globally optimal solution CVs, it is assumed that the NCO is approximated by CVs and the CVs are perfectly controlled at zero. Then, the loss, $L(d)$ for a particular disturbances d, due to the approximation error, $\epsilon(d)$ can be expressed as [1], [9]:

$$L(d) = 0.5\epsilon(d)^T H(d)^{-1} \epsilon(d) \quad (12)$$

where $H(d)$ is the reduced Hessian of the cost function evaluated at point where the CV, $c(d)$ is perfectly controlled corresponding to particular disturbance, d, whilst $\epsilon(d) = g(d) - c(d)$, where $g(d)$ is the reduced gradient evaluated at the same point.

The average loss over the entire operation region, D can be represented as,

$$\bar{L} = E_{d\in D} L(d) \approx \frac{1}{2N} \sum_{i=1}^{N} \epsilon(d_i)^T H(d_i)^{-1} \epsilon(d_i) \quad (13)$$

where $d_i \in D$, $i = 1, \ldots, N$ are N samples of disturbances in D and E is the expectation operator.

According to (13), the loss minimization is equivalent to a least squares regression problem to minimize the weighted residual, $H^{-1/2}\epsilon$. However, due to the difficulty and reliability to evaluate the reduced Hessian for every $d_i \in D$, $H(d_i)$ can be replaced by a constant matrix, *e.g.* the identity matrix or H evaluated at nominal value of d. Then the regression problem can be set up as discussed next.

Let the entire operation region be sampled by N points for independent variables (input and disturbance), $u = \begin{bmatrix} u_1 & \cdots & u_N \end{bmatrix}$ and $d = \begin{bmatrix} d_1 & \cdots & d_N \end{bmatrix}$. The corresponding measurement values and the reduced gradient values are $y = \begin{bmatrix} y_1 & \cdots & y_N \end{bmatrix}$ and $g = \begin{bmatrix} g_1 & \cdots & g_N \end{bmatrix}$, respectively. The CV is parameterized by θ at each sampling point as:

$$c_i = f_\theta(y_i),\ i = 1, \ldots, N \tag{14}$$

where $f_\theta(\cdot)$ is the parameterized regression function of measurements, which can be either linear or nonlinear, such as polynomial or Gaussian. Then the optimal CV, $c^* = f_{\theta^*}(y)$ is determined by adjusting θ to minimize the regression error $\epsilon = f_\theta(y_i) - g_i$ as follows:

$$\min_\theta \ \frac{1}{2N} \sum_{i=1}^{N} (f_\theta(y_i) - g_i)^T H^{-1} (f_\theta(y_i) - g_i) \tag{15}$$

For linear regression, $f_\theta(y) = \theta y$, *i.e.* $g = \theta y + \epsilon$. Then the regression problem in (15) can be solved analytically; see Section IV-B for details. In principle, it is possible to parametrize the CVs in terms of all the available measurements. Control systems with lower complexity can be obtained by using a subset of available measurements to parametrize the CVs, which often provides similar loss as the case where CVs are chosen to be functions of all the available measurements [6], [5], [7], [9]. The selection of the subset of measurements is a combinatorial optimization problem, which makes the application of this method difficult to large-scale processes. The BAB framework used to overcome this difficulty is presented in the next section.

IV. Branch and bound method

A. General principle

Let $X_r = \{x_i | i = 1, 2, \cdots, r\}$, be an r-element set. A subset selection problem with the selection criterion ϕ involves finding the optimal solution, X_n^*, such that

$$\phi(X_n^*) = \max_{X_n \subset X_r} \phi(X_n) \tag{16}$$

For this problem, the number of alternatives is $\mathcal{C}_r^n = \frac{r!}{(r-n)!n!}$, which grows very quickly with r and n rendering exhaustive search unviable. A BAB approach can provide globally optimal solution for the subset selection problem in (16) without exhaustive search. In this approach, the original problem (node) is divided (branched) into several non-overlapping subproblems (sub-nodes). If any of the n-element solution of a sub-problem cannot lead to the optimal solution, the sub-problem is not evaluated further (pruned), else it is branched again. The pruning of sub-problems allows the BAB approach to gain efficiency in comparison with exhaustive search.

The available BAB methods for subset selection can be classified as downwards [12], [14], [15], [16], [17] and upwards [18], [10], [11] BAB methods based on the search direction. For the regression problem associated with globally optimal CV selection, the downwards BAB approach is applicable and is discussed next. The reader is referred to [12], [14], [17] for details on downwards BAB method.

In a downwards BAB approach, each node is represented by $X_s = F_f \cup C_c$, where $s = f + c$ and, F_f and C_c denote the fixed and candidate sets, respectively. Here, the subscript denote the size of the set. The relationship between the fixed and candidate sets of a node and its i^{th} sub-node (branching rule) is given as follows:

$$F_{f_i}^i = F_f \cup \{x_1, \cdots, x_{i-1}\};\ C_{c_i}^i = C_c \setminus \{x_1, \cdots, x_i\} \tag{17}$$

where $F_{f_i}^i$ and $C_{c_i}^i$ denote the fixed and candidate sets of the i^{th} sub-node and $i = 1, 2, \cdots, n - f + 1$. An example of the solution tree obtained by recursively applying the branching rule in (17) is shown in Figure 1. For the root node in this solution tree, we have $F_f = \emptyset$ and $C_c = X_r$. The label of the nodes denote the element being removed from X_s. The solution tree has $\mathcal{C}_n^r$ terminal nodes, which represent different n-element subsets of X_r.

To describe the pruning principle, let $\mathcal{X}$ denote the ensemble of all n-element subsets, which can be obtained using (17), *i.e.*

$$\mathcal{X} = \{F_f \cup C_c / X_{f+c-n} | X_{f+c-n} \in C_c\} \tag{18}$$

and $\underline{\phi}(F_f \cup C_c)$ be the upper bound on ϕ computed over all elements of $\mathcal{X}$, *i.e.*

$$\underline{\phi}(F_f \cup C_c) = \max_{X_n \in \mathcal{X}} \phi(X_n) \tag{19}$$

Assume that B is a lower bound of the globally optimal criterion, *i.e.* $B \leq \phi(X_n^*)$. Then,

$$\phi(X_n) < \phi(X_n^*) \quad \forall X_n \in \mathcal{X}, \quad \text{if } \underline{\phi}(F_f \cup C_c) < B \tag{20}$$

Hence, any $X_n \in \mathcal{X}$ cannot be optimal and can be pruned without further evaluation, if $\underline{\phi}(F_f \cup C_c) < B$.

Although pruning of nodes using (20) results in an efficient BAB algorithm, further efficiency can be gained by performing pruning on the sub-nodes directly. This happens as the lower bounds for different sub-nodes are related and can be computed together from $\underline{\phi}(F_f \cup C_c)$ resulting in computational efficiency. For $x_i \in C_c$, the i^{th} sub-node can be pruned if

$$\underline{\phi}(F_f \cup C_c / x_i) < B \tag{21}$$

For a BAB method involving pruning of sub-nodes, branching needs to be carried on sub-node level as well, which requires choosing a decision element to branch upon. Here, the decision element is selected as the element with largest $\underline{\phi}(F_f \cup C_c / x_i)$ among all $x_i \in C_c$ (best-first search).

B. Application to CV Selection using Regression

The linear regression model is given as:

$$b = A\theta + \epsilon \tag{22}$$

where b are the observations or measurements, matrix A contains the regressors, θ are the unknown parameters and ϵ is the noise. Under the assumption that ϵ is independently

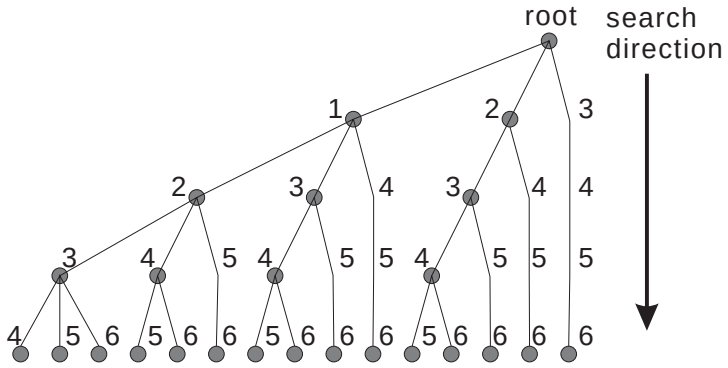

Fig. 1. Solution tree for selecting 2 out of 6 elements

and identically distributed (i.i.d.), it is well known that the unbiased estimate of θ is given as; see *e.g.* [19],

$$\hat{\theta} = \left(A^T A\right)^{-1} A^T b \tag{23}$$

With the estimate of θ given in (23), the predicted values of observations are $\hat{b} = A\hat{\theta} = A\left(A^T A\right)^{-1} A^T b$ and the prediction error e is given as

$$e = b - \hat{b} = Pb \tag{24}$$

where $P = \left(I - A\left(A^T A\right)^{-1} A^T\right)$. Then, sums of squares of errors (SSE) can be computed as

$$SSE = e^T e = b^T P^T P b = b^T P b \tag{25}$$

where the last identity follows as P is an idempotent matrix [19]. The SSE can be further expressed as

$$SSE = b^T b - b^T A\left(A^T A\right)^{-1} A^T b \tag{26}$$

As the first term in (26) is constant, the variables can be selected by minimizing SSE or equivalently maximizing $b^T A\left(A^T A\right)^{-1} A^T b$. Let us define $y = A^T b$ and $C = A^T A$. Now, the subset selection can be performed by solving the following optimization problem:

$$\max_{X_n \subset X_r} \phi(X_n) = \mathbf{y}_{X_n}^T (\mathbf{C}_{X_n,X_n})^{-1} \mathbf{y}_{X_n} \tag{27}$$

where $X_r = \{1, 2, \cdots, r\}$, $\mathbf{y}_{X_n}$ denotes the elements of $\mathbf{y}$ with indices in X_n and $\mathbf{C}_{X_n,X_n}$ represents the principal submatrix of $\mathbf{C}$ with rows and columns indexed by X_n. Note that a similar combinatorial optimization problem is considered in [20], where the objective function instead needs to be minimized.

The use of BAB for solving the optimization problem in (27) requires an upper bound on the selection criteria, calculated over the ensemble $\mathcal{X}$ in (18). This upper bound is derived in the next proposition.

***Proposition** 1:* Consider a node with fixed set F_f and candidate set C_c. For $\mathcal{X}$ in (18),

$$\phi(F_f \cup C_c) \geq \max_{X_n \in \mathcal{X}} \phi(X_n) \tag{28}$$

Proposition 1 implies that the non-optimal nodes can be pruned using $\phi(F_f \cup C_c)$ as the upper bound. To gain further efficiency by pruning the sub-nodes directly, we relate the selection criteria of a node with its sub-nodes in the next proposition.

***Proposition** 2:* Consider a node with fixed set F_f and candidate set C_c. Let $S = F_f \cup C_c$. For $x_i \in C_c$, $i = 1, 2, \cdots, c$,

$$\phi(S \setminus x_i) = \phi(S) - \alpha_i^2/\delta_i \tag{29}$$

where

$$\alpha_i = \mathbf{z}_i^T \mathbf{y}_S \tag{30}$$

whilst $\mathbf{z}_i^T$ and δ_i are the ith row and (i,i)th element of $\mathbf{C}_{S,S}^{-1}$ respectively.

The evaluation of (29) requires inversion of only one matrix $\mathbf{C}_{S,S}$, which is the same for all $x_i \in C_c$. Thus, the use of (29) to obtain the selection criteria for all sub-nodes together is computationally more efficient than directly evaluating the selection criteria for every node. In summary, the following BAB algorithm can be used for subset selection for regression.

***Algorithm** 1:* Initialize $f = 0$, $F_f = \emptyset$, $C_c = X_r$, $\phi(F_f) = 0$ and $B = 0$. Call the following recursive algorithm:

1) If $\phi(F_f \cup C_c) > B$, prune the current node and return, else perform the following steps.
2) Calculate α_i in (30) $\forall i \in C_c$. Prune the subsets with $\phi(F_f \cup C_c) - \alpha_i^2 < B$.
3) If $f = n$ or $f + c = n$, go to next step. Otherwise, generate the c sub-nodes according to the branching rule in (17) and call the recursive algorithm in Step 1 for each sub-node. Return to the caller after the execution of the loop finishes.
4) Find $J_{\max} = \phi(F_f \cup C_c) - \max_{i \in C_c} \alpha_i^2$. If $J_{\max} > B$, update $B = J_{\max}$. Return to the caller.

V. Binary Distillation Column Case Study

To evaluate the efficiency of proposed BAB algorithm for selecting globally optimal CVs, we test the performance of BAB algorithm on a binary distillation column [13]. All tests are conducted on a PC running Windows 7 SP1 with Intel Core i3-2100 3.10GHz processor, 8GB RAM using Matlab R2011a.

The objective is to minimize the relative steady-state deviations of the distillate ($z^L_{\rm top}$) and bottoms ($z^H_{\rm btm}$) compositions from their nominal values, *i.e.*

$$J = \left(\frac{z^H_{\rm top} - z^H_{\rm top,s}}{z^H_{\rm top,s}} \right)^2 + \left(\frac{z^L_{\rm btm} - z^L_{\rm btm,s}}{z^L_{\rm btm,s}} \right)^2 \quad (31)$$

where the superscripts L and H refer to the light and heavy components and the nominal steady-state values are $z^H_{\rm top,s} = z^L_{\rm btm,s} = 0.01$ (99% purity). The distillation column has 4 manipulated variables: reflux flow rate (L), vapor boilup (V), distillate flow rate (D) and bottoms flow rate (B). Because the levels of top condenser and bottom reboiler need to be stabilized, which consumes two degrees of freedom. We select D and B to control the levels, which is also referred as LV configuration for distillation column control, therefore, two degrees of freedom are remained for composition control. The main disturbances are feed flow rate (F), feed composition (z_F) and vapor fraction of feed (q_F), which are allowed to vary between 1 ± 0.2, 0.5 ± 0.1 and 1 ± 0.1, respectively. The top and bottom compositions are not measured online and thus two CV's needs to be identified for indirect control of the compositions. It is considered that the temperatures on 41 trays ($y_1, \ldots, y_{41}$, counting from bottom to top) are measured with an accuracy of $\pm 0.5^\circ$C, whose combinations can be used as CVs for implementation of self-optimizing control strategy.

Data samples for NCO regression are generated as follows: each independent variable is evenly divided into 5 parts within its variation range. The variation range for disturbances are defined earlier. Variation range for reflux flow rate L is chosen as $1 \pm 10\%$ at its nominal value and vapor boilup V is bounded within $(L-(1-q_F)F, L+q_F F)$ in order to let $0 < B$,$D < F$. For each scenario, temperatures at each tray are calculated and the two NCO components J_L and J_V, which refer to the gradient of J with respect to L and V, respectively, are also obtained using finite difference method. Therefore, $6^5 = 7776$ samples are collected for regression. Because the number of candidate measurements for regression is large, we apply proposed BAB algorithm to choose an appropriate subset and determine globally optimal CVs. Similar computation performances are observed for regressing J_L and J_V, as summarized in Figure 2.

Figure 2 (a) and (b) show that using full set of measurements as predictors for NCOs is not necessary. When $n > 5$, the SSE can only be slightly reduced. A trade-off between SSE, which is directly related to the overall economic loss, and the number of measurements used has to be made. Therefore, we can choose $n = 6$ and get CV models for self-optimizing control of this column. Figure 2 (c), (d), (e) and (f) show the computation time and number of node evaluations and demonstrate the usefulness and effectiveness of proposed BAB algorithm. Brute force cannot handle such a large problem, whereas proposed BAB algorithm solves it successfully. It takes about 1260s to complete all the selection tasks, which is reasonable for off-line computing. Largest computation time is seen for $n = 15$, which takes about 100s for selecting 15 out of 41 measurements. Overall, proposed algorithm is practically appealing, as the algorithm makes it possible to reduce the overall operation cost and meanwhile, reduces the investment for hardware sensors (*e.g.* temperature sensors for the column).

Furthermore, the economic losses associated with the CVs obtained by using the proposed BAB algorithm are evaluated and compared with those associated with CVs obtained in previous work [11] through local self-optimizing control. It is found that $n = 6$ results in a good trade-off between self-optimizing performance and investment cost of sensors, where both CVs are designed using 6 measurements.

The self-optimizing method using local average criterion [5] is adopted for comparison. To derive the CVs, linear model and Hessian matrices at the nominal point are obtained through finite difference. We also choose $n = 6$ and apply the PB^3 algorithm [11], [22] to get the best measurement subset.

A Monte Carlo experiment with 100 sets of randomly generated disturbances within their allowable ranges is used. The feedback control actions are implemented to maintain CVs at 0 and objective function J is calculated, whose optimal value is expected to be 0. The average and maximal objective costs for regression method are 2.222 and 82.766, respectively, while the average and maximal objective costs for local method are 170.411 and 1577.474, respectively. This is because that the distillation process is strongly nonlinear, therefore, local method fails to characterize the process in wider operation range, whereas the regression method is advantageous in this respect resulting in significantly enhanced self-optimizing performance.

VI. Conclusions

In the context of self-optimizing control, a novel branch and bound (BAB) algorithm is proposed for selecting globally optimal controlled variables (CVs) based on the method of regression for necessary conditions of optimality. The BAB algorithm aims to identify the best measurement set to minimize the sums of squares of errors (SSE) and use the set to construct the optimal CVs. Numerical tests using a practical binary distillation column case study show the efciency and effectiveness of this algorithm. It is pointed out that the proposed algorithm is applicable to the general linear regression problem as well as other statistical problems. The proposed downwards algorithm is efficient for problems where a few among many candidate variables need to be discarded. However, for problems where a few variables needs to be selected from many candidate variables, the computational expense incurred by the algorithm is still large. To this end, upwards and bidirectional BAB algorithms are currently being developed and will be presented in an extended paper.

References

[1] L. Ye, Y. Cao, Y. Li, and Z. Song, "A data-driven approach for selecting controlled variables," in *to be presented at the 8th Intl. Symposium on ADCHEM*, Singapore, 2012.

[2] S. Skogestad, "Plantwide control: The search for the self-optimizing control structure." *J. Proc. Control*, vol. 10, no. 5, pp. 487–507, 2000.

[3] G. P. Rangaiah and V. Kariwala, *Plantwide Control: Recent Developments and Applications*. Chichester, UK: John Wiley & Sons, 2012.

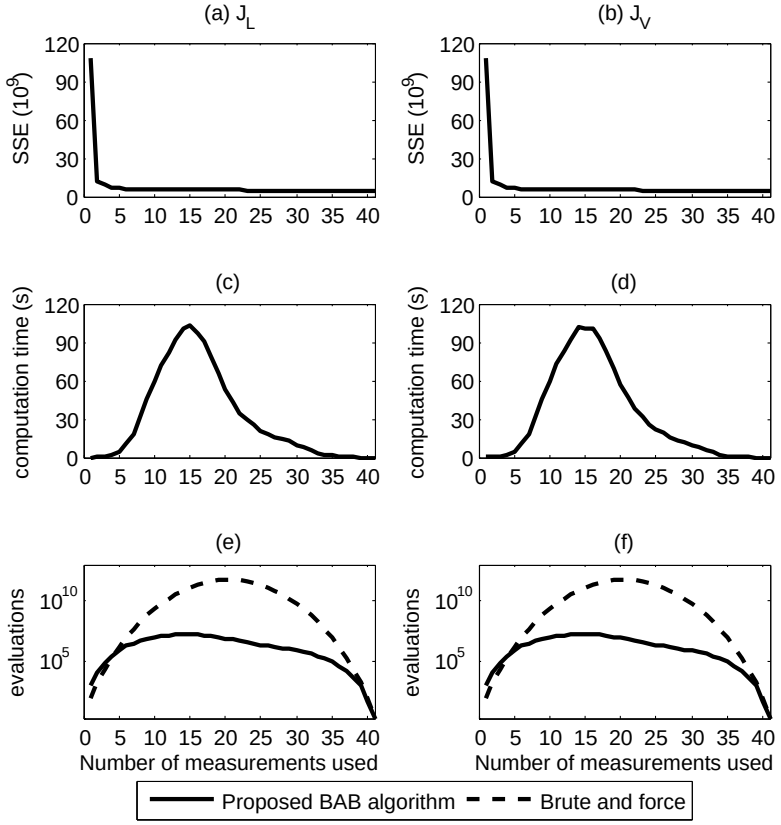

Fig. 2. BAB performance for the column case study: (a) regression SEE for J_L; (b) regression SEE for J_V; (c) computation time for J_L; (b) computation time for J_V; (e) Evaluations for J_L; (f) Evaluations for J_V

[4] I. J. Halvorsen, S. Skogestad, J. C. Morud, and V. Alstad, "Optimal selection of controlled variables," *Ind. Eng. Chem. Res.*, vol. 42, no. 14, pp. 3273–3284, 2003.

[5] V. Kariwala, Y. Cao, and S. Janardhanan, "Local self-optimizing control with average loss minimization," *Ind. Eng. Chem. Res.*, vol. 47, no. 4, pp. 1150–1158, 2008.

[6] V. Kariwala, "Optimal measurement combination for local self-optimizing control," *Ind. Eng. Chem. Res.*, vol. 46, no. 11, pp. 3629–3634, 2007.

[7] V. Alstad, S. Skogestad, and E. S. Hori, "Optimal measurement combinations as controlled variables," *J. Proc. Control*, vol. 19, no. 1, pp. 138–148, 2009.

[8] W. Hu, L. M. Umar, G. Xiao, and V. Kariwala, "Local self-optimizing control of constrained processes," *J. Proc. Contr.*, vol. 22, pp. 488–493, 2012.

[9] L. Ye, Y. Cao, Y. Li, and Z. Song, "Approximating necessary conditions of optimality as controlled variables," *Ind. Eng. Chem. Res.*, p. submitted, 2012.

[10] V. Kariwala and Y. Cao, "Bidirectional branch and bound for controlled variable selection: Part II. Exact local method for self-optimizing control," *Comput. Chem. Eng.*, vol. 33, no. 8, pp. 1402–1412, 2009.

[11] ——, "Bidirectional branch and bound for controlled variable selection: Part III. Local average loss minimization," *IEEE Trans. Ind. Informat.*, vol. 6, no. 1, pp. 54–61, 2010.

[12] P. Narendra and K. Fukunaga, "A branch and bound algorithm for feature subset selection," *IEEE Trans. Comput.*, vol. C-26, pp. 917–922, 1977.

[13] S. Skogestad, "Dynamics and control of distillation columns - A tutorial introduction," *Trans. IChemE Part A*, vol. 75, pp. 539–562, 1997.

[14] B. Yu and B. Yuan, "A more efficient branch and bound algorithm for feature selection," *Pattern Recognition*, vol. 26, pp. 883–889, 1993.

[15] P. Somol, P. Pudil, F. Ferri, and J. Kittler, "Fast branch and bound algorithm in feature selection," in *Proceedings of World Multiconference on Systemics, Cybernetics and Informatics*, B. Sanchez, J. Pineda, J. Wolfmann, Z. Bellahsense, and F. Ferri, Eds., vol. VII, Orlando, Florida, USA, 2000, pp. 1646–651.

[16] X.-W. Chen, "An improved branch and bound algorithm for feature selection," *Pattern Recognition Letters*, vol. 24, pp. 1925–1933, 2003.

[17] Y. Cao and P. Saha, "Improved branch and bound method for control structure screening," *Chem. Engg. Sci.*, vol. 60, no. 6, pp. 1555–1564, 2005.

[18] Y. Cao and V. Kariwala, "Bidirectional branch and bound for controlled variable selection: Part I. Principles and minimum singular value criterion," *Comput. Chem. Engng.*, vol. 32, no. 10, pp. 2306–2319, 2008.

[19] R. Johnson and D. Wichern, *Applied multivariate statistical analysis*. Prentice hall Upper Saddle River, NJ, 2002, vol. 4.

[20] V. Kariwala, P. Odiowei, Y. Cao, and T. Chen, "A branch and bound method for isolation of faulty variables through missing variable analysis," *Journal of Process Control*, vol. 20, no. 10, pp. 1198–1206, 2010.

[21] Y. Cao, D. Rossiter, and D. H. Owens, "Output effectiveness and scaling sensitivity for secondary measurement selection," *Transaction of IChemE, Part A*, vol. 76, pp. 849–854, 1998.

[22] Y. Cao and V. Kariwala, "B3AV," MATLAB File Exchange, November 2009, available at http://www.mathworks.com/matlabcentral/fileexchange/25870.