Proceedings of the 17th World Congress
The International Federation of Automatic Control
Seoul, Korea, July 6-11, 2008

Iterative Feedback Tuning of Cross-directional Processes Controller ⋆

Jun Yan * **Guy Dumont** ** **Philip D. Loewen** ***

* *Pulp and Paper Center, University of British Columbia, Vancouver, BC V6T 1Z4, Canada; (e-mail: juny@ece.ubc.ca).*
** *Pulp and Paper Center, University of British Columbia, Vancouver, BC V6T 1Z4, Canada; (e-mail: guyd@ece.ubc.ca)*
*** *Department of Mathematics, University of British Columbia, Vancouver, BC V6T 1Z2, Canada; (e-mail: loew@math.ubc.ca)*

Abstract: This paper studies the possible application of Iterative Feedback Tuning (IFT) to the paper machine cross-directional (CD) control. Although the CD control is naturally a multi-input-multi-output (MIMO) problem, by enforcing the *circulant* assumption, IFT for CD control is as simple as the single-input-single-output (SISO) case, i.e. only one gradient experiment is needed for estimating the gradient. Simulation of a simple disturbance rejection problem is included to demonstrate the main ideas.

Keywords:
Data-based control, Control of particulate processes, Autotuning

1. INTRODUCTION

This paper is part of a study on automatic tuning of paper machine cross-directional (CD) control [VanAntwerp et al., 2007], a rather special multi-input-multi-output (MIMO) control problem. The main goals of CD control are to maintain uniform CD paper properties at the set grade, to reject process disturbances, and to save control energy (or hardware from wearing). These goals can be formulated into an L_2 minimization problem. With many successful simple-structured CD controllers working in industry, Iterative Feedback Tuning (IFT) comes naturally into consideration for further performance improvement.

The idea of IFT [Hjalmarsson et al., 1998] is to minimize an L_2 cost function $J(y, u)$ in terms of process output and control signals, with respect to the controller parameters ρ via a numerical optimization algorithm, e.g. gradient-descent methods [Gill et al., 1986]. As the exact functional form of $J(\rho)$ is unknown, the necessary quantities (usually the gradient and hessian) for calculating numerical search directions are estimated based on experimental data extracted from the real plant. The data needed for one parameter update/iteration consist of a set of normal outputs and some gradient experiment outputs and do not require plant downtime. Because $J(\rho)$ is typically a nonlinear function, the algorithm very likely converges to a local minimum.

IFT is often referred as *data-based* and *model-free*, although the line between *model-free* and *model-based* becomes blurred when the IFT algorithm is performed based on spectrum estimates [Kammer et al., 2000]. One may argue that using fresh data to perform a model-update then a model-based controller redesign. As discussed in [Hjalmarsson et al., 1995], in the case of low-complexity fixed-structured controllers, IFT shows advantages in improving performance without increasing the controller complexity. If high-complexity controllers were allowed, then a model-based update should be more desirable.

IFT is reported very effective in tuning SISO systems [Hjalmarsson et al., 1998]. Despite the number of parameters, only one set of gradient experiment is need for an SISO system. As for MIMO case, the number of experiments (hence time!) required to finish one update may become quite large. As shown in [Hjalmarsson, 1999], for a system with n_u inputs and n_y outputs, the required number of gradient experiments is $\min(\#\textit{nonzero control channels}, \#\rho)$. For the disturbance rejection only problem, the first element may decrease to $n_u + n_y$.

Paper-making is a $2D$ process, i.e. the paper properties depend on both space and time. A paper machine uses several actuator arrays, at the Headbox, the press section, and the dryer section, to control several paper properties, such as caliper, basis-weight, and moisture. A typical single-actuator-array to single-property model takes the MIMO form: the output vector consists of the samples along the CD direction (or spatial index), the input vector represents the actuators in an array mounted along the CD direction, the transfer matrix between them is the product of a constant spatial interaction matrix and a diagonal transfer matrix whose elements are the same temporal transfer function. The spatial interaction matrix is usually a *Toeplitz* matrix characterized by only a few parameters, usually much less than the dimension of the matrix.

In this paper, we propose using IFT to tune a square single-actuator-array to single-property CD controller. An industrial CD controller may involve $30 \sim 300$ actuators in

⋆ This work was supported in part by the National Sciences and Engineering Research Council of Canada.

one array and $200 \sim 2000$ output samples of one property. The CD controller of interest is used in more than 4200 paper machines worldwide. It is often initially tuned via the *two-dimensional frequency-domain loop-shaping* [Stewart et al., 2003], which delivers a working controller with enough robust stability margin and provides a good initial point for an IFT scheme that may improve the performance further.

At first look, the paper machine CD control should fall into the MIMO case in which a gradient experiment should be performed for each parameter. Recall that the simplicity in the SISO case arises from the commutative property of the scalar transfer functions in the output gradient calculation, which is usually invalid in the MIMO case. However, it is a common assumption in CD controller design that the spatial interaction matrices are *circulant* as used in [Stewart et al., 2003]. According to [Gray], circulant matrices commute and their sums and inverses are still circulant. Hence, the corresponding output gradient calculation can be simplified similarly to the SISO case and only one set of gradient experiment is needed for each update. The circulant assumption can be enforced by including appropriate compensation at the edges. This idea is studied numerically in a simple but informative disturbance rejection problem.

2. A BRIEF REVIEW OF ITERATIVE FEEDBACK TUNING

Suppose the plant of interest is described by

$$\begin{aligned} Y(t) &= GU(t) + w(t), \\ U(t) &= K(\rho)(R(t) - Y(t)), \end{aligned} \tag{1}$$

where $Y(t) \in R^n$ is the plant output vector, $U(t) \in R^m$ is the control vector, $w(t) \in R^n$ is the process output disturbance, $R(t) \in R^n$ is the reference signal, G is the unknown linear transfer function matrix, and $K(\rho)$ is the linear feedback control law which is characterized by parameter vector $\rho \in R^p$ with ρ_i being the ith element.

The goal of IFT is to improve the performance via minimizing the following cost function

$$\begin{aligned} J(\rho) = \frac{1}{2N} E \sum_{j=1}^{N} \{ & (Y(j) - Y^d(j))^T (Y(j) - Y^d(j)) \\ & + \lambda U^T(j) U(j) \}, \end{aligned} \tag{2}$$

where the tuning variables are the controller parameters ρ and the desired performance/behavior of the plant is specified via $Y^d(t) = T^d R(t)$. The cost function $J(\rho)$ is typically non-convex in ρ. If the exact function $J(\rho)$ were known, then a gradient-based numerical algorithm such as the Newton-Raphson method

$$\rho[k+1] = \rho[k] - \gamma_k H^{-1}[k] \frac{\partial J}{\partial \rho} \mid_{\rho=\rho[k]}, \tag{3}$$

would bring the sequence $\{\rho[k]\}$ to at least a local minimum of $J(\rho)$. In (3), the matrix H is the associated Hessian matrix.

However, due to the involvement of the unknown plant G in J, the calculation of its gradient is impossible. Hence the key idea of IFT is to estimate the gradient from the real outputs and the algorithm is changed to

$$\rho[k+1] = \rho[k] - \gamma_k R^{-1}[k] \widehat{\frac{\partial J}{\partial \rho}} \mid_{\rho=\rho[k]}, \tag{4}$$

where $R[k]$ is a positive definite matrix and $R^{-1}[k] \widehat{\frac{\partial J}{\partial \rho}} \mid_{\rho=\rho[k]}$ is still a descent direction. Usually $R[k]$ is chosen to be the estimated Hessian

$$R[k] = \widehat{\frac{\partial J}{\partial \rho}} \widehat{\frac{\partial J}{\partial \rho}}^T \mid_{\rho=\rho[k]}. \tag{5}$$

The gradient of J with respect to ρ_i (or the ith element of the gradient $\frac{\partial J}{\partial \rho}$) is

$$\frac{\partial J}{\partial \rho_i} = \frac{1}{N} E \sum_{j=1}^{N} \{ (Y(j) - Y^d(j))^T \frac{\partial Y}{\partial \rho_i} + \lambda U^T(j) \frac{\partial U}{\partial \rho_i} \}, \tag{6}$$

$$\frac{\partial Y}{\partial \rho_i} = -S_0^2 G \frac{\partial K}{\partial \rho_i} GKR + S_0 G \frac{\partial K}{\partial \rho_i} R - S_0^2 G \frac{\partial K}{\partial \rho_i} w, \tag{7}$$

$$\frac{\partial U}{\partial \rho_i} = \frac{\partial K}{\partial \rho_i}(Y - R) + K \frac{\partial Y}{\partial \rho_i}, \tag{8}$$

where $S_0 = (1 + GK)^{-1}$ and $T_0 = (1 + GK)^{-1} GK$ are the achieved sensitivity and the complementary sensitivity respectively. It is evident from (6) $\sim$ (8) that the core issue of estimating the gradient is to estimate $\frac{\partial Y}{\partial \rho_i}$.

If the plant is a SISO system, the gradient (7) can be greatly simplified since the transfer matrices become scalar transfer functions that commute:

$$\frac{\partial Y}{\partial \rho_i} = \frac{\partial K}{\partial \rho_i} K^{-1} T_0 (R - T_0 R - S_0 w) = \frac{\partial K}{\partial \rho_i} K^{-1} T_0 (R - Y). \tag{9}$$

Based on (9), the following three experiments should generate data that are necessary for estimating $\frac{\partial Y}{\partial \rho_i}, i = 1 \ldots p$,

(1) $R^1 = R$, $Y^1 = T_0 R + S_0 w^1$;
(2) $R^2 = R + F(R - Y^1)$, $Y^2 = T_0 R + T_0 F(R - Y^1) + S_0 w^2$;
(3) $R^3 = R$, $Y^3 = T_0 R + S_0 w^3$.

In the above proposed experiments, a signal with superscript $i = 1, 2, 3$ means an N-sample realization of that signal, the controller is parameterized by the current $\rho[k]$, and F is a pre-filter. These three experiments are similar to the ones described in [Hjalmarsson et al., 1998], except the second one using $R^2 = R + F(R - Y^1)$ instead of $R = (R - Y^1)$, in the hope that most of the products (Y^2) generated in the second experiment are still acceptable. The output gradient can be estimated as follows:

$$\begin{aligned} \widehat{\frac{\partial Y}{\partial \rho_i}} &= \frac{1}{N} \sum_{j=1}^{N} \{ \frac{\partial K}{\partial \rho_i} K^{-1} F^{-1} (Y^3(j) - Y^2(j)) \}, \\ &= \frac{1}{N} \sum_{j=1}^{N} \{ \frac{\partial K}{\partial \rho_i} K^{-1} T_0 (R(j) - Y^1(j)) \\ &+ \frac{\partial K}{\partial \rho_i} K^{-1} F^{-1} S_0 (w^3(j) - w^2(j)) \}. \end{aligned} \tag{10}$$

Note that the pre-filter F should be designed so that *i*) the noise term $\frac{\partial K}{\partial \rho_i} K^{-1} F^{-1} S_0 (w^3(j) - w^2(j))$ is insignificant in the gradient estimate (10), which requires F to be a high-pass filter, and *ii*) most of Y^2 can still be accepted as products. The estimate (10) is unbiased if the noises w^2 and w^3 are zero-mean and independent.

For the MIMO case, as shown in [Hjalmarsson, 1999], the gradient experiment needs to inject $\frac{\partial K}{\partial \rho_i}(R^1 - Y^1)$ into the closed-loop system as the input disturbance and the N-sample gradient experiment should be repeated $\min(p, mn)$ times in order to generate $\frac{\partial J}{\partial \rho_i}$ for all $i = 1, \ldots, p$. Thus the time required to finish one update of the parameters in the MIMO case may be inadmissible to many and become the limiting factor to apply IFT to MIMO systems.

Once the estimated output gradient $\widehat{\frac{\partial Y}{\partial \rho}}$ is obtained, the estimated gradient of the cost function can be obtained by

$$\widehat{\frac{\partial J}{\partial \rho_i}} = \frac{1}{N}\sum_{j=1}^{N}\{(Y^1(j) - Y^d(j))^T \widehat{\frac{\partial Y}{\partial \rho_i}} + \lambda U^T(j)\widehat{\frac{\partial U}{\partial \rho_i}}\}, \tag{11}$$

$$\widehat{\frac{\partial U}{\partial \rho_i}} = \frac{1}{N}E\sum_{j=1}^{N}\{\frac{\partial K}{\partial \rho_i}(Y^1 - R^1) + K\widehat{\frac{\partial Y}{\partial \rho_i}}\}. \tag{12}$$

Then the update direction should follow:

$$\Delta\rho[k] = R^{-1}[k]\widehat{\frac{\partial J}{\partial \rho_i}}, \tag{13}$$

with $R^{-1}[k]$ being the estimated Hessian (5). The new parameters are obtained by (4):

$$\rho[k+1] = \rho[k] - \gamma_k R^{-1}[k]\widehat{\frac{\partial J}{\partial \rho}}\ |_{\rho=\rho[k]}\ .$$

In numerical optimization context [Gill et al., 1986], choosing the proper γ_k is known as the line search problem for speedy convergence of the algorithm. However, in IFT, a more important question is whether $K(\rho[k+1])$ will be a stabilizing controller — the problem of *cautious tuning*. For SISO systems, the ν-gap measure between $K(\rho[k])$ and $K(\rho[k+1])$ may provide certain guidance to determine $\rho[k+1]$ as proposed in [Kammer et al., 2000] and a relaxed version is proposed in [Kammer, 2005].

3. IFT FOR PAPER MACHINE CD CONTROLLER

In the paper machine CD control context, the model is developed under the standard assumptions that (1) the paper-making process is linear, temporally and spatially invariant; (2) the process response is separable, i.e. it can be factored into a temporal-only response times a spatial-only response; (3) the spatial response is symmetric.

A typical paper machine model describing one actuator array to one paper property is

$$Y(t) = \frac{1}{1 - a_0 z^{-1}} z^{-d} BU(t) + w(t), \tag{14}$$

where the output $Y \in R^{N_a}$ is the CD-profile of the paper property, the input $U \in R^{N_a}$ is the CD-profile of the actuator array, the machine direction (dynamic) model is a first-order element with a pole at a_0 and a pure delay element z^{-d}. The spatial interaction matrix $B = Toeplitz(\bar{b}, N_a)$ is a $N_a \times N_a$ Toeplitz matrix generated from a vector $\bar{b}$, where N_a is the number of actuators and the spatial response is captured by the vector $\bar{b} = [b_{l_b}, \ldots, b_1, b_0, b_1, \ldots, b_{l_b}]$, $l_b \ll N_a$. Although (14) is an empirical model, it will be used as the real plant in our IFT study.

The CD controller of interest $K(z)$ is defined as follows

$$\begin{aligned} K(z) &= [1 - Dz^{-1}]^{-1}D \cdot c(z)I_{N_a} \cdot C, \qquad (15)\\ C &= Toeplitz(\bar{c}, N_a), \bar{c} = [c_{l_c}, \ldots, c_0, \ldots, c_{l_c}], l_c \ll N_a,\\ c(z) &= \frac{(1-\alpha)(1 - a_0 z^{-1})}{(1 - a_0)[1 + (1-\alpha)\sum_{i=1}^{d-1} z^{-i}]}, \alpha \in [0, 1],\\ D &= Toeplitz(\bar{d}, N_a), \bar{d} = [d_{l_d}, \ldots, d_0, \ldots, d_{l_d}], l_d \ll N_a. \end{aligned}$$

This CD controller consists of three parts: (1) the spatial decoupling filter C; (2) the Dahlin controller $c(z)$; (3) the actuator profile smoother $[1 - Dz^{-1}]^{-1}D$. The initial tuning of this controller is usually done via $2D$ loop-shaping [Stewart et al., 2003], where the model parameters a_0, d are used directly in the Dahlin controller part and the main tuning parameters are $\bar{c}$, α, and $\bar{d}$.

If the CD controller (15) were to be tuned via IFT, it would fall into the MIMO case discussed in Section 2. This means between the two normal experiments, we ought to run $size(\bar{c}) + size(\bar{d}) + 1$ gradient experiments if we were to tune $\bar{c}$, α, and $\bar{d}$ as $l_c,\ l_d \ll N_a$. However, the paper machine measuring mechanism, the transversing scanner [VanAntwerp et al., 2007], is typically a slow process producing $2 \sim 3$ CD-profiles per minute. This means it may take days to collect the necessary data to finish one update of the parameters while assuming the plant is producing the same grade of paper 24/7. This scenario makes IFT much less appealing.

Note that, the simplicity in the SISO case arises from reducing (7) to (9) via exercising the commutative property of the scalar transfer functions under multiplication. Although, in general, the transfer matrices do not commute (neither do the Toeplitz matrices B, C, and D), a common *circulant* approximation can lift this hindrance on the CD control tuning. The *circulant* approximation was used in [Stewart et al., 2003] to greatly simplify the $2D$ loop-shaping process, in which the Toeplitz matrices B, C, and D were replaced by circulant matrices generated by the same vectors. In other words, the process is thought to be producing “paper tube” instead of “paper sheet”. In the real paper making process, there exist compensating actuators beyond the edges described in the model (14) and (15) to reduce the edge effect. Hence the *circulant* approximation may be valid if the compensating actuators for one edge are set to be the reflection of the actuators near the opposite edge. Then the following lemma can guarantee that the reduction from (7) to (9) works for the paper machine CD control.

Lemma 1. (from [Gray]). If B and C are circulant $n \times n$ matrices, then

(1) B and C commute;
(2) $B + C$ is a circulant matrix;
(3) for nonsingular B, B^{-1} is a circulant matrix.

Therefore, the SISO IFT scheme described in Section 2 should work for the paper machine CD controller with circulant spatial interaction matrices.

For the tuning of the spatial elements $\bar{c}$ and $\bar{d}$, non-causal transfer functions [Ammar and Dumont, 2005] can be used so that the tuning of $\bar{c}$ and $\bar{d}$ becomes equivalent to the tuning of the parameters in the corresponding non-causal spatial filters. For example, if $\bar{c}$ is the impulse response of the following noncausal filter $C(q, q^{-1})$

$$C_0(q) = \frac{e_0 q^{n_c} + e_1 q^{n_c - 1} + \cdots + e_{n_c}}{q^{n_c} + f_1 q^{n_c - 1} + \cdots + f_{n_c}},$$
$$C(q, q^{-1}) = C_0(q) + C_0(q^{-1}), \tag{16}$$

where q and q^{-1} are the right- and left- shift operators in the spatial index, then the tuning of $\bar{c}$ is equivalent to the tuning of $\{e_0, \ldots, e_{n_c}, f_1, \ldots, f_{n_c}\}$.

The use of the non-causal model in IFT may bring us the following advantages:

(1) The vector $\bar{c}$ is subject to a structural constraint — its length l_c is fixed. Thus the tuning may reduce the length of $\bar{c}$, but it is not clear when to extend it. The non-causal model does not suffer this limitation.
(2) Similar to temporal industrial controllers, where low-order filters can produce good performance, if we entrust first- and second-order $C_0(q)$ to product good enough performance, then only a small number of parameters need to be tuned and the necessary calculation load to generate gradient estimates is also reduced.
(3) The robust stability result using ν-gap measure developed in [Ammar and Dumont, 2005] may be extended to provide cautious spatial controller tuning criteria that are similar to the temporal case as in [Kammer et al., 2000] and [Kammer, 2005].

4. SIMULATION STUDY

In this Section, we apply the proposed IFT to a simple paper machine disturbance rejection problem. The model used as the plant is

$$Y(t) = \frac{1}{1 - a_0 z^{-1}} z^{-d} B U(t) + w(t),$$

where the number of CD actuators and measurements is $N_a = 54$, the process pole is $a_0 = 0.8311$, and the delay is $d = 3$. The spatial interaction matrix B is the circulant matrix generated by

$$\begin{aligned}\bar{b} =& [0.0713, 0.0337, -0.0167, -0.02, -0.005, 0.0006,\\ & 0.0005, 0.0001].\end{aligned}$$

The disturbance $w(t)$ has normal distribution $N(0, I_{54})$. Except for $w(t)$, the model corresponds to a paper mill producing 'telephone book grade' paper.

Usually the cost function (2) can be regarded as the sum of three costs: tracking error, disturbance rejection, and control energy. Here as the first attempt of applying IFT on paper machine control problem, we consider only the disturbance rejection part by setting $R(t) = Y^d \equiv 0$ and $\lambda = 0$. Therefore the cost function becomes

$$J(\rho) = \frac{1}{2N} E\left\{ \sum_{j=1}^{N} Y^T(j) Y(j) \right\}. \tag{17}$$

As the disturbance is $w(t) \sim N(0, I_{54})$, the optimal parameters are simply: $\alpha = 1$ or $\bar{c} = 0$ or $\bar{d} = 0$ and the optimal cost is $cov(w(t)) = 54$, i.e. by shutting down the controller we have the output behave just like the uncontrollable white noise $w(t)$. Here we consider the tuning of $\bar{c}$ only. Since $R = 0$, only two experiments are needed:

(1) $R^1 = 0$, $Y^1 = S_0 w^1$;
(2) $R^2 = -Y^1$, $Y^2 = -T_0 Y^1 + S_0 w^2$;

where no pre-filter F is considered and the gradient estimates are calculated by

$$\widehat{\frac{\partial Y}{\partial \rho_i}} = \frac{1}{N} \sum_{j=1}^{N} \{ \frac{\partial K}{\partial \rho_i} K^{-1}(Y^2(j)) \},$$
$$\widehat{\frac{\partial J}{\partial \rho_i}} = \frac{1}{N} \sum_{j=1}^{N} \{ Y^1(j)^T \widehat{\frac{\partial Y}{\partial \rho_i}} \}.$$

With this simple, though unrealistic problem, we can demonstrate certain features of applying IFT to paper machine.

4.1 Tuning of spatial interaction matrices

We first study tuning the matrix C in the controller $K(z)$ which is defined in (15)

$$\begin{aligned} K(z) &= [1 - D z^{-1}]^{-1} D \cdot c(z) I_{54} \cdot C,\\ C &= Circulant(\bar{c}, 54), \bar{c} = [c_{l_c}, \ldots, c_0, \ldots, c_{l_c}],\\ c(z) &= \frac{(1-\alpha)(1 - a_0 z^{-1})}{(1 - a_0)[1 + (1-\alpha)\sum_{i=1}^{d-1} z^{-i}]}, a \in [0, 1],\\ D &= Circulant(\bar{d}, 54), \bar{d} = [d_{l_d}, \ldots, d_0, \ldots, d_{l_d}].\end{aligned}$$

The initial parameters are

$$\begin{aligned}&\alpha = 0.96, \{d_0, \ldots, d_3\} = \{0.986, 0.0046, 0.002, 0.0004\},\\ &\{c_0, \ldots, c_5\} = \{10.4708, 3.5297, -1.4841, 0.0042,\\ &\qquad -0.0017, -0.0006\}.\end{aligned}$$

And here we assume that the dynamical part (a_0 and d) of the plant is known. The tuning process was stopped

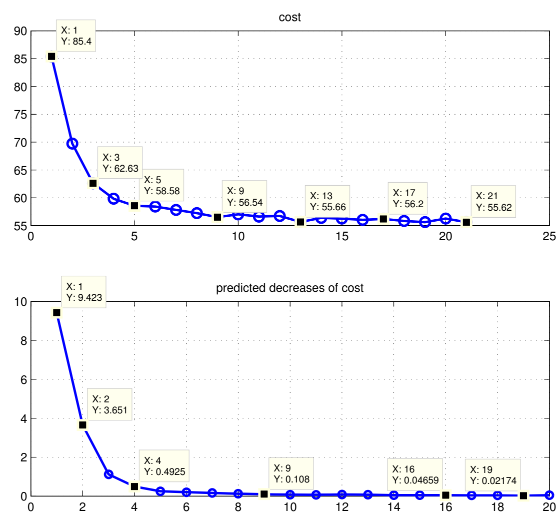

Fig. 1. The cost and the predicted decreases in cost via tuning $\bar{c}$.

after 20 iterations. In each iteration 2000 CD-profiles were measured. The step size γ_k in the parameters update (4) is set to 1. Figure 1 shows the cost function values and the predicted decreases of the cost against the number of iterations. The cost was reduced from 85.40 to 55.62 and the final $\bar{c}$ is $[4.6346, -0.8876, -0.1951, 0.9509, -0.2574, 0.0417]$.

The reduction in the cost was significant during the first a few iterations. After that, the predicted change of cost

$$\Delta J[k] = \rho[k]\widehat{\frac{\partial J}{\partial \rho}},$$

is close to zero, where $\rho[k]$ is $\bar{c}$ in the kth iteration. The simulation shows that small $\Delta J[k]$ agrees with near-zero gradient estimate of the cost function $\widehat{\frac{\partial J}{\partial \rho}}$. In general, near-zero $\widehat{\frac{\partial J}{\partial \rho}}$ should indicate one of the following situations:

(1) the controller parameters are close to the global optimal ones, or;
(2) the current controller parameters in the loop are near a local minimum but not the global one, or;
(3) the signal-to-noise ratio in the gradient experiment is very low so that the gradient estimate is mainly calculated from the measured noises.

Further improvement is possible for the latter two cases. To increase the low signal-to-noise ratio, we may include a proper pre-filter F in the gradient experiment. The local minimum case could be viewed as the current controller parameters $\rho[k]$ being a 'bad' initial point for $\min_\rho J(\rho)$ via the gradient-descent algorithm (4), if it were treated as a numerical optimization problem. However, as $J(\rho)$ is unknown, it is impossible to guess a good initial point. More importantly, abrupt change in the controller parameters in the loop can be very dangerous to the plant. Hence to keep the algorithm going, changing $\rho[k]$ cannot be the driving force, we should change the desired behavior T^d instead — we may regard the current situation as the desired behavior T^d in the cost function being, in a sense, far away from the achieved behavior T_0. Therefore, using a less demanding T_1^d may fix the problem and the original T^d can be injected back later. The specific how-to's are left for future study. Figure 2 shows the distributions of

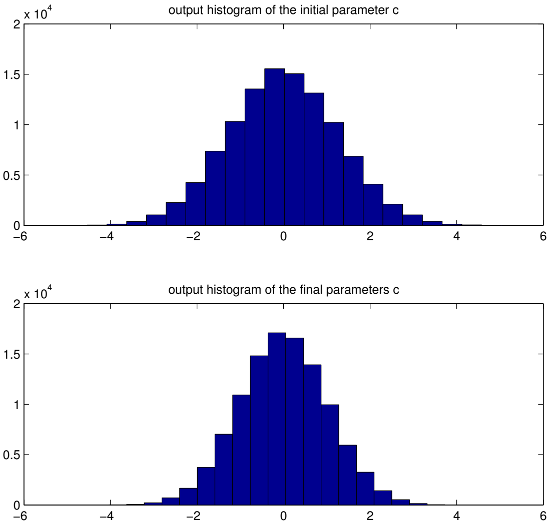

Fig. 2. The distributions of the output corresponding to the initial and final $\bar{c}$.

the measured outputs corresponding to the initial and final $\bar{c}$, where the final distribution is more concentrated.

4.2 Tuning of the non-causal model

As discussed in Section 3, the vector $\bar{c}$ is equivalent to the impulse response of a non-causal filter. Here we consider using a second-order non-causal filter:

$$\begin{aligned} C_0(q) &= \frac{e_0 q^2 + e_1 q + e_2}{q^2 + f_1 q + f_2}, \\ C(q, q^{-1}) &= C_0(q) + C_0(q^{-1}). \end{aligned} \tag{18}$$

The tuning parameters are $\bar{e} = [e_0, e_1, e_2]$ and $\bar{f} = [f_1, f_2]$ with initial values $\bar{e} = [10.4700, 3.5310, -1.4960]$ and $\bar{f} = [0.0001, -0.0011]$ that generate the same initial $\bar{c}$ in Section 4.1. The simulation was done by using the spatial matrices configuration (15), while $\bar{c}$ was taken to be the impulse response of (18) whose small magnitude ($< 10^{-4}$) elements were set to zero. The simulation results Figure 3 and Figure 4 are very similar to those in Section 4.1.

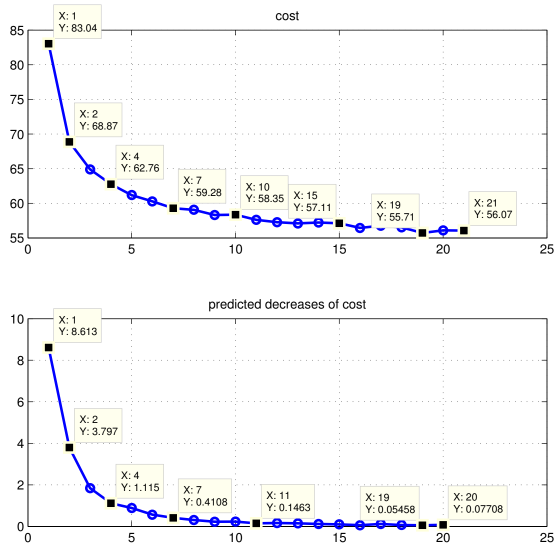

Fig. 3. The cost and the predicted decreases in cost via tuning $\bar{e}$ and $\bar{f}$.

5. CONCLUSIONS AND FUTURE WORKS

We have shown that IFT is a good candidate of automatic tuning schemes for paper machine CD controller. By enforcing the circulant approximation, IFT for single-actuator-array to single-property CD control can be greatly simplified like IFT for SISO systems. Future studies may include:

- running simulation to include more realistic disturbance rejection and tracking problems (although the latter is less important for CD control);
- extending the ν-gap measure to $2D$ controllers (spatial and temporal) that provides guide to stable parameter update for CD control;
- developing a concrete IFT scheme tailored for paper machine CD controllers tuning that includes pre-filter design, parameter update criteria, stopping criteria, etc;

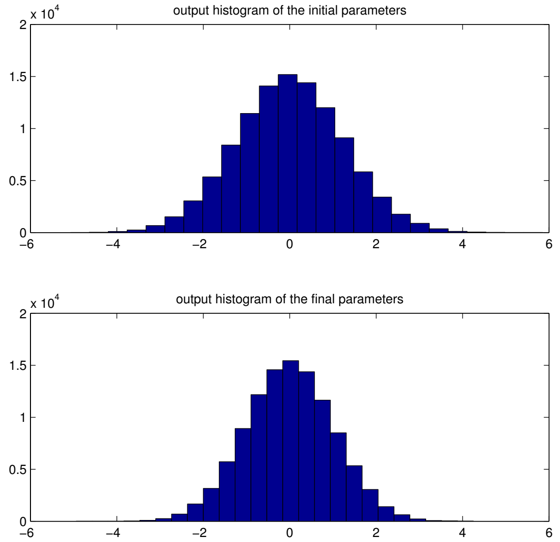

Fig. 4. The distributions of the output corresponding to the initial and final $\bar{e}$ and $\bar{f}$.

– exploring the possibility of IFT for multiple-actuator-array to multiple-property CD control that utilizes the existing MIMO IFT results.

REFERENCES

M. Ammar and G. Dumont. Robust stability analysis of cross-directional processes using the ν-gap metric. In *Proceedings of the 45th IEEE Conference on Decision & Control*, pages 1203–1208, San Diego, CA, 2005.

P.E. Gill, W. Murray, and M.H. Wright. *Practical Optimization*. Elsevier & Ltd., 1986.

R.M. Gray. Toeplitz and circulant matrices: A review. URL `http://ee.stanford.edu/ gray/toeplitz.pdf`.

H. Hjalmarsson. Effecient tuning of linear multivariable controllers using iterative feedback tuning. *International Journal of Adaptive Control and Signal Processing*, 13: 553–572, 1999.

H. Hjalmarsson, S. Gunnarsson, and M. Gevers. Optimality and sub-optimality of iterative identification and control design schemes. In *Proc. American Control Conf.*, volume 4, pages 2559–2563, Seattle, Washington, 1995.

H. Hjalmarsson, M. Gevers, S. Gunnarsson, and O. Lequin. Iterative feedback tuning: Theory and applications. *IEEE Control Systems*, 18:26–41, 1998.

L.C. Kammer. Stability assessment for cautious iterative controller tuning. *Automatica*, 41:1829–1834, 2005.

L.C. Kammer, R.R. Bitmead, and P.L. Bartlett. Direct iterative tuning via spectral analysis. *Automatica*, 36: 1301–1307, 2000.

G.E. Stewart, D.M. Gorinevsky, and G.A. Dumont. Two-dimensional loop shaping. *Automatica*, 39:779–792, 2003.

J.G. VanAntwerp, A.P. Featherstone, R.D. Braatz, and B.A. Ogunnaike. Cross-directional control of sheet and film processes. *Automatica*, 43:191–207, 2007.