SIAM J. Matrix Anal. Appl.
Vol. 19, No. 3, pp. 664–681, July 1998

# CHOOSING POLES SO THAT THE SINGLE-INPUT POLE PLACEMENT PROBLEM IS WELL CONDITIONED*

VOLKER MEHRMANN† AND HONGGUO XU‡

**Abstract.** We discuss the single-input pole placement problem (SIPP) and analyze how the conditioning of the problem can be estimated and improved if the poles are allowed to vary in specific regions in the complex plane. Under certain assumptions we give formulas as well as bounds for the norm of the feedback gain and the condition number of the closed loop matrix. Via several numerical examples we demonstrate how these results can be used to estimate the condition number of a given SIPP problem and also demonstrate how to select the poles to improve the conditioning.

**Key words.** pole placement, condition number, perturbation theory, Jordan form, Cauchy matrix, stabilization, feedback gain, distance to uncontrollability, optimal conditioning

**AMS subject classifications.** 65F15, 65F35, 65G05, 93B05, 93B55

**PII.** S0895479896302382

## 1. Introduction.

We consider linear single-input control systems

$$\dot{x} := dx/dt = Ax + bu, \qquad x(0) = x_0, \tag{1}$$

where $A \in \mathcal{C}^{n\times n}$, $b \in \mathcal{C}^n$, $x$, $u$ are functions defined on $[0, +\infty) \rightarrow \mathcal{C}^n$ and $[0, +\infty) \rightarrow \mathcal{C}$, respectively. For such systems, we consider the single-input pole placement problem.

**Single-input pole placement (SIPP).** *For a given set of poles $\mathcal{P} = \{\lambda_1, \ldots, \lambda_n\}$, find a feedback gain vector $f \in \mathcal{C}^n$ such that the set of eigenvalues of the closed-loop matrix $A - bf^T$ is $\mathcal{P}$.*

(Note that we use $f^T$, although $f$ may be complex, since this simplifies the formulas.)

It is well known that for an arbitrary pole set $\mathcal{P}$, the feedback $f$ always exists and is unique if and only if $(A, b)$ is controllable, see, e.g., ([20, page 48, Theorem 2.1]). This problem has been studied extensively. In the literature there are some explicit formulas known for $f$ and the Jordan canonical form of $A - bf^T$ as well as many perturbation results; see [1, 11, 12, 2, 17]. Also, many numerical algorithms have been proposed for this problem; see [2, 4, 10, 14, 15, 16, 18, 8]. In order to analyze the validity of the computed results and the robustness of the solution, it is very important to study the sensitivity of the solution with respect to perturbations in the data. This question has lead to some confusion in the literature [10, 11, 8, 12]. This confusion arises mainly from the fact that there are essentially two different types of results for the SIPP, namely, the feedback gain vector $f$ and the eigenstructure of the closed loop matrix $A - bf^T$. It is possible and actually quite common that, even though $f$ is insensitive to perturbations in the data, the spectrum of $A - bf^T$ is very

*Received by the editors April 22, 1996; accepted for publication (in revised form) by P. Van Dooren August 21, 1997; published electronically March 18, 1998.
http://www.siam.org/journals/simax/19-3/30238.html

†Fakultät für Mathematik, TU Chemnitz, D-09107 Chemnitz, Germany (mehrmann@mathematik.tu-chemnitz.de). The research of this author was partially supported by DFG project Me 790/7-2: Singuläre Steuerungsprobleme.

‡Department of Mathematics, Fudan University, Shanghai 200433, People's Republic of China. Current address: Fakultät für Mathematik, TU Chemnitz, D-09107 Chemnitz, Germany (hxu@mathematik.tu-chemnitz.de). The research of this author was supported by the Alexander von Humboldt Foundation and the Chinese National Natural Foundation.

sensitive, and vice versa. Examples 1 and 4 below demonstrate this phenomenon. It follows that there are also two condition numbers to study here, one for the mapping from $(A, b, \mathcal{P})$ to $f$ and one for the mapping $(A, b, \mathcal{P})$ to the eigenstructure of $A - bf^T$. This observation, and the fact that it is often not explicitely stated which solution of the SIPP problem is considered, explains some of the confusion in the literature. From the point of view of applications, in our opinion the more important problem is to guarantee that the poles of the computed closed loop system $A - bf^T$ are close to the desired ones; the accuracy of $f$ is less important. In particular, in applications such as stabilization it would be fatal if the closed loop poles were made unstable by very small perturbations. To see that this may happen very easily and unexpectedly, consider the following example.

*Example* 1 (see [12]). Consider the SIPP problem with data

$$A = \operatorname{diag}(1, 2, \ldots, 15), \quad b = \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix}, \quad \mathcal{P} = \{-1, \ldots, -15\}.$$

In this case, both $f$ and $A - bf^T$ can be computed analytically and hence no rounding errors occur in these quantities; see [12] for details. But the eigenvalues of the closed loop systems are so sensitive to perturbations that some of the computed eigenvalues of $A - bf^T$ are in the right half plane.

Unfortunately, this example is not an exception, as was pointed out in [8] and partially proved in [12]; the SIPP problem is usually ill-conditioned. This means that in most cases, in particular if the system size is large ($n > 10$), small perturbations in the data $A, b, \mathcal{P}$ will cause large perturbations in the eigenstructure of $A - bf^T$. In practice it can be expected that if such a feedback is implemented, then the real behavior of the closed loop system is very different than the expected behavior.

Based on these results we have to reconsider the pole placement problem. Since in applications one almost never needs the poles of the closed-loop system in fixed positions but rather in specific regions in the complex plane, it is natural to ask the question of whether we can optimize the conditioning of the problem by varying the choice of poles in the prescribed regions. If we reconsider pole placement in this way, then we have the following problem.

**Optimal single-input pole placement (OSIPP).** *Given $A \in \mathcal{C}^{n \times n}$, $b \in \mathcal{C}^n$ and a set $\mathcal{D} \subset \mathcal{C}$, find poles $\lambda_1, \ldots, \lambda_n \in \mathcal{D}$, i.e., $\mathcal{P} \subset \mathcal{D}$, such that the SIPP problem is optimally conditioned among all possible choices of $\mathcal{P} \subset \mathcal{D}$.*

To study this problem we first have to discuss what the condition number of the SIPP problem is and how it can be computed or estimated. In particular it would be important for an optimization to have an easily computable quantity or estimate.

In general the condition number of a problem measures the sensitivity of the solution with repect to perturbations in the input data. As we have mentioned above, several quantities can be viewed as solutions of the pole placement problem and different formulas and bounds have been given in recent years; see [2, 8, 11, 17, 12].

We will base our optimization on modifications of the formula for the condition number given in [17]. This formula is very difficult to compute so that an optimization for this condition number seems hopeless. In [12], therefore, based on explicit solution formulas for $f$ and the closed loop eigenvector matrix, slightly different bounds were obtained. We will modify these bounds again to obtain quantities which we can optimize.

To do this we introduce the following notation. The *scaled spectral condition number* of a diagonalizable matrix $A$ is defined as

$$\kappa := \|G\|\,\|G^{-1}\|,$$

where $G$ is the eigenvector matrix of $A$ normalized such that all columns have unit norm. The scaled spectral condition number is equivalent to the optimal spectral condition number; see [5]. Here we use $\|\cdot\|$ to denote an arbitrary consistent norm and use $\|\cdot\|_2$, $\|\cdot\|_F$ to denote the Euclidean and Frobenius norm, respectively. We denote by $\kappa$, $\kappa_2$, and $\kappa_F$ the associated scaled spectral condition numbers of $A-bf^T$. By $\Lambda(A)$ we denote the set of eigenvalues of a square matrix $A$, and by $\sigma_1(B) \geq \cdots \geq \sigma_p(B) \geq 0$, $p = \min\{m,n\}$ we denote the singular values of an $m \times n$ matrix $B$; see [7]. By $\mathcal{C}_0^+$, $\mathcal{C}^-$, and $\mathcal{C}_{-\rho}^-$ we denote closed right half plane, open left half plane, and the set of complex numbers with real parts not larger than $-\rho$ for $\rho > 0$, respectively. Finally we set $e$ to be the vector of all ones and $e_i$ to be the $i$th unit vector.

The following perturbation theorem is a combination of two perturbation results given in [12], with slightly modified assumptions.

THEOREM 1.1. *Consider the SIPP problem with data $A, b, \mathcal{P} = \{\lambda_1, \ldots, \lambda_n\}$. Assume that $(A, b)$ is controllable and that the $n$ poles $\lambda_i$ are distinct. Let $\lambda = [\lambda_1, \ldots, \lambda_n]^T$. Consider also the perturbed problem with data $\hat{A} := A + \delta A$, $\hat{b} := b + \delta b$, and $\delta\lambda = [\delta\lambda_1, \ldots, \delta\lambda_n]^T$. Assume that $(\hat{A}, \hat{b})$ is also controllable and that also the perturbed poles are distinct. Set $\epsilon := \max\{\|[\delta A \ \ \delta b]\|, \max_i |\delta\lambda_i|\}$ and suppose that*

$$2\epsilon < \min_i \sigma_n \begin{bmatrix} A - \lambda_i I & b \end{bmatrix} =: \sigma_\lambda.$$

*Let $f$, $\hat{f} := f + \delta f$ be the feedback gains of the original and perturbed problems, respectively, then*

$$\|\delta f\|_2 \leq c_f := \frac{2\sqrt{2n}}{\sigma_\lambda}\epsilon\hat{\kappa}_2\sqrt{1+\|f\|_2^2}\,\max_i \sqrt{\left(\frac{\|\hat{A}-\hat{\lambda}_i I\|_2}{\|\hat{b}\|_2}\right)^2 + 1}, \tag{2}$$

*where $\hat{\kappa}_2$ is the scaled spectral condition number of $\hat{A} - \hat{b}\hat{f}^T$.*

*Furthermore, for each eigenvalue $\mu_i$ of the computed closed-loop matrix $A - b\hat{f}^T$, there exists a corresponding eigenvalue $\lambda_i$ of the desired (unperturbed) closed-loop system such that*

$$|\mu_i - \lambda_i| \leq c_e := \left(1 + \hat{\kappa}_2\sqrt{1+\|\hat{f}\|_2^2}\right)\epsilon. \tag{3}$$

*Proof.* The proof is easily obtained from the proofs of Theorems 7 and 8 in [12], using the slightly different assumptions on $\delta A$, $\delta b$, $\delta\lambda$. The estimates (2), (3) are obtained analogously. □

We see from Theorem 1.1 that several factors contribute to the perturbation bounds and thus can be considered to create large perturbations in $f$ and/or the closed-loop eigenvalues. The main factors in the bound (2) are the quantity $\sigma_\lambda$ and the term $\hat{\kappa}\sqrt{1+\|f\|_2^2}$. In the following we restrict our optimization to the second factor. This may lead to an overestimation of the bound in the optimal case but it simplifies the optimization and can be justified as follows. The term $\sigma_\lambda$ reflects the distance to uncontrollability $d_{uc}(A, b) = \min_{s\in\mathcal{C}} \sigma_n[A - sI, b]$, which is independent

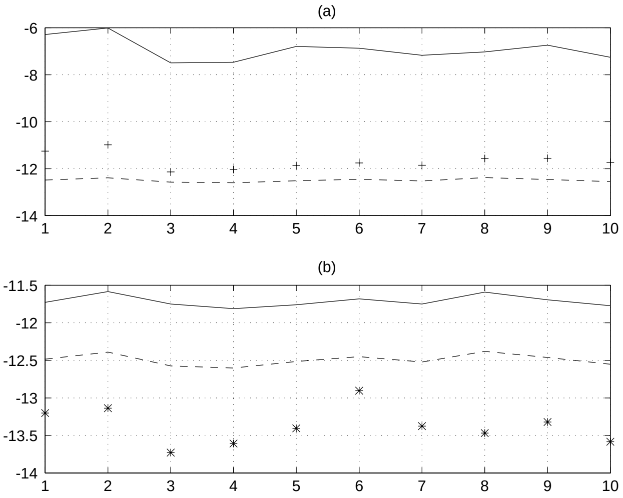

FIG. 1. *Error and bounds in Example* 2: (a) *feedback gain:* $\log(\|f-\hat{f}\|) : +++$, $\log(c_f)$ : —, $\log(eps \cdot \mathcal{S})$ : - - -; (b) *closed-loop poles*: $\log(eig_e) : ***$, $\log(c_e)$ : —, $\log(eps \cdot \mathcal{S})$ : - - -.

of the choice of poles. So, we can replace $\sigma_\lambda$ in (2) to obtain an upper bound. But it should be noted that $d_{uc}(A,b)$ can be much smaller than $\sigma_\lambda$, see, e.g., [3]. If $d_{uc}(A,b)$ is reasonably large, then replacing $\sigma_\lambda$ by $d_{uc}(A,b)$ will have only a small effect on the bound. If, however, $d_{uc}(A,B)$ is very small, then this may lead to an overestimation of the condition number. This effect is demonstrated in some of our numerical examples below. But in this case we know that the problem is very close to a problem which is not controllable. This is a critical situation in practice and it may be reasonable to modify the model in such a case. Consider now the term $\hat{\kappa}\sqrt{1+\|f\|_2^2}$, which governs the perturbations in the computed closed-loop eigenvalues. This term (if large) also makes the bound (2) very large. If this term can be made small by the choice of poles and if $d_{uc}(A,b)$ is reasonably large, then both bounds (2) and (3) are small and we can expect that $f$ can be computed accurately and that the closed loop eigenvalues are robust. We will therefore optimize the quantity

$$\mathcal{S} := \kappa\sqrt{1+\|f\|^2} \tag{4}$$

to improve the conditioning via the choice of poles. If necessary, we use the notation $\mathcal{S}_2 := \kappa_2\sqrt{1+\|f\|_2^2}$ and $\mathcal{S}_F := \kappa_F\sqrt{1+\|f\|_2^2}$. In order to illustrate that $\mathcal{S}$ catches the qualitative behavior of the errors well, we will consider the following numerical tests.

All computations in this paper were carried out in Matlab Version 4.2 on a pentium-s PC with machine precision $eps = 2.22 \times 10^{-16}$. Random matrices are created with the Matlab *rand* function and uniform distribution. In the figures below we depict $c_f$ as in (2), $c_e$ as in (3) with $\epsilon = \max\{\|[A\ b]\|, \max_i |\lambda_i|\}eps$, and by $eig_e$ we denote the maximal error between an eigenvalue of the computed closed-loop matrix and the associated pole in $\mathcal{P}$.

*Example* 2. In this example we constructed ten problems with $A \in \mathcal{R}^{30\times 30}$, $b \in \mathcal{R}^{30}$ with elements chosen randomly in $[-1,1]$, and for each of these problems we chose 50 different (exact) feedback gains $f \in \mathcal{R}^{30}$ with random elements in $[-10,10]$. For each of these 500 problems the chosen poles are the computed eigenvalues of $A - bf^T$ (via the Matlab *eig* function). Then we computed the feedback gain $\hat{f}$ with

TABLE 1
min *and* max *of* $\log(c_f/\|f-\hat{f}\|)$.

| Problem | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|---|---|
| min | 3.6 | 3.3 | 3.6 | 3.3 | 3.6 | 3.8 | 3.4 | 3.4 | 3.6 | 3.1 |
| max | 6.5 | 7.0 | 5.9 | 6.1 | 7.1 | 6.8 | 6.4 | 5.9 | 6.0 | 5.7 |

TABLE 2
min *and* max *of* $\log(c_e/eig_e)$.

| Problem | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|---|---|
| min | -0.1 | -0.1 | 1.6 | 1.2 | 1.3 | 0.9 | 1.0 | 1.1 | 0.9 | 0.8 |
| max | 3.2 | 3.1 | 3.2 | 2.9 | 2.1 | 3.1 | 2.8 | 3.3 | 2.8 | 2.6 |

these poles by using Miminis and Paige's *sevas* Matlab code (cf. [14]). In Figure 1 for each of the 10 pairs $(A, b)$ we display the arithmetic means of logarithms of $c_e$, $eig_e$, $c_f$ and $\|f-\hat{f}\|$ taken over the 50 experiments and compare it with $eps\cdot\mathcal{S}$. We also display the minimum and maximum distances between the orders of $c_f$ and $\|f-\hat{f}\|$, $e_f$, and $eig_e$ in Tables 1 and 2, respectively, among the 50 experiments for each pair $(A, b)$.

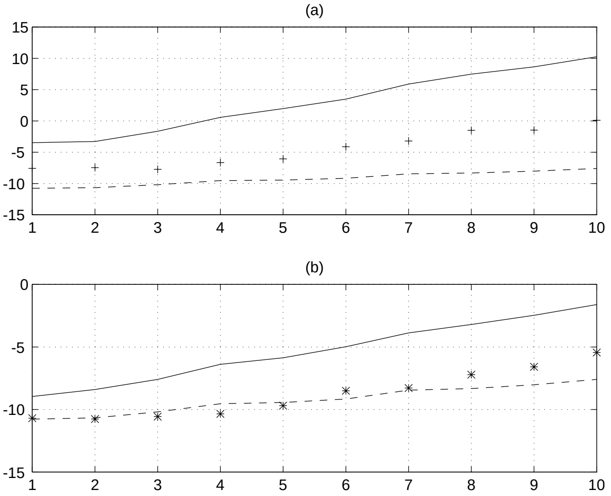

FIG. 2. *Error and bounds in Example* 3: (a) *feedback gain:* $\log(\|f-\hat{f}\|) : +++$, $\log(c_f)$ : —, $\log(eps\cdot\mathcal{S})$ : - - -; (b) *closed-loop poles:* $\log(eig_e) : ***$, $\log(c_e)$ : —, $\log(eps\cdot\mathcal{S})$ : - - -.

We see that the bounds and also the term $\mathcal{S}$ describe the qualitative behavior of the errors quite well, although the bounds (2) and (3) sometimes tend to be too pessimistic. In general we cannot expect to see more than the qualitative behavior, since we have omitted terms in the bounds. In Example 2 we have that $\sigma_\lambda$ contributes a factor $10^2$, which almost explains the difference of the bounds.

The estimate of the errors in the closed-loop eigenvalues is in general better than that of the feedback gain, since the factors do not occur.

The reason for the negative numbers in Table 2 is an underestimation of $\epsilon$.

*Example* 3. The conditioning becomes worse if we modify Example 2 by diagonal scaling of $A$. Let $A := D\tilde{A}D^{-1}$, $D = \operatorname{diag}(1, 2^{\frac{m+2}{3}}, \ldots, 30^{\frac{m+2}{3}})$ with $\tilde{A}$ as in Example 2. Here we let $m$ take the values 1 to 10, and the poles are formed as before. The

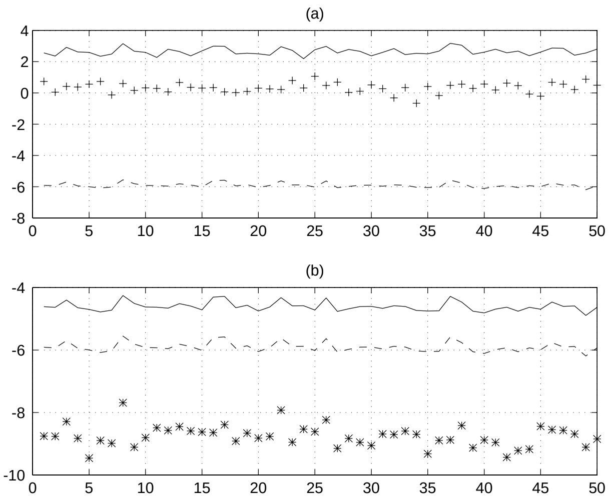

FIG. 3. *Error and bounds in Example* 4: (a) *feedback gain:* $\log(\|f - \hat{f}\|) : + + +$, $\log(c_f) :$ —, $\log(eps \cdot \mathcal{S}) :$ - - -; (b) *closed-loop poles:* $\log(eig_e) : * * *$, $\log(c_e) :$ —, $\log(eps \cdot \mathcal{S}) :$ - - -.

errors and bounds for this example are in given Figure 2.

Our next example demonstrates that it is not sufficient to consider the accuracy of the feedback gain $f$ alone. In this case the computed closed-loop eigenvalues are much more accurate than the computed $f$.

*Example* 4. Let

$$A = Q^T \begin{bmatrix} -1 & -1 & \cdots & -1 \\ 1 & -1 & \ddots & \vdots \\ \mathbf{O} & \ddots & \ddots & \vdots \\ & & 1 & -1 \end{bmatrix} Q \in \mathcal{R}^{30\times 30}, \quad b = Q^T e_1 \in \mathcal{R}^{30},$$

where $Q$ is a random orthogonal matrix, $f = 10^5 Q^T f_1$, and $f_1 \in \mathcal{R}^{30}$ has elements randomly selected in $[-1, 1]$. We chose 50 random vectors $f_1$ and produced the poles as in Example 2.

Figure 3 shows that, even though on the average the computed $f$ has no correct digit, the closed loop eigenvalues still carry essentially eight correct digits.

## 2. Minimization of $\mathcal{S}$.

In view of the discussion in the previous section, we now consider the minimization of $\mathcal{S}$, as defined in (4), when the closed-loop eigenvalues are allowed to vary in a given set $\mathcal{D} \subset \mathcal{C}$.

In the following we restrict our minimization to the case that all elements in $\mathcal{P}$ are distinct so that $A - bf^T$ is diagonalizable. Although condition numbers are also interesting in the degenerate case, they are more complicated and usually the conditioning is much worse ([19, pages 87–90]).

If we consider the SIPP problem with data $A, b, \lambda$, where $(A, b)$ is controllable and the components of $\lambda$ are distinct, then explicit formulas for the solution of the SIPP problem are known; see [1, 4, 12]. For the case of distinct poles we have the following formula from [12]. Let

$$G = [u_1, \ldots, u_n], \quad \|u_i\|_2 = 1, \quad i = 1, \ldots, n, \tag{5}$$

and $\alpha = [\alpha_1, \dots, \alpha_n]$ such that for each $i$, $[u_i^T, -\alpha_i]^T$ is nonzero and satisfies

$$[A - \lambda_i I, b] \begin{bmatrix} u_i \\ -\alpha_i \end{bmatrix} = 0. \tag{6}$$

Then $G$ is nonsingular,

$$f^T = \alpha^T G^{-1} = e_n^T [b, Ab, \dots, A^{n-1}b]^{-1} \prod_{i=1}^{n} (A - \lambda_i I), \tag{7}$$

$$A - bf^T = G \operatorname{diag}(\lambda_1, \dots, \lambda_n) G^{-1},$$

and

$$\kappa = \|G\| \, \|G^{-1}\|. \tag{8}$$

In principle we could employ nonlinear optimization methods to compute the required minimum, but considering the explicit formulas (5), (6), and (7), this is a very difficult problem, in particular when $n$ is large.

We will now discuss some cases where we can give explicit formulas for $\|f\|$ and $\kappa$ and thus, also, the optimization problem becomes much simpler.

THEOREM 2.1. *Let $A = \Gamma := \operatorname{diag}(\gamma_1, \dots, \gamma_n)$, $b = e$, and $(\Gamma, e)$ be controllable, let $\mathcal{P} = \{\lambda_1, \dots, \lambda_n\}$ be a pole set with distinct elements and $\Lambda(A) \cap \mathcal{P} = \emptyset$. Then*

$$\|f\|_2^2 = \sum_{i=1}^{n} \frac{\prod_{k=1}^{n} |\gamma_i - \lambda_k|^2}{\prod_{k=1, k\neq i}^{n} |\gamma_i - \gamma_k|^2}, \tag{9}$$

$$\kappa_F^2 = n \sum_{i,j=1}^{n} \frac{\sum_{l=1}^{n} \prod_{k=1,k\neq l}^{n} |\lambda_i - \gamma_k|^2}{\prod_{k=1,k\neq i}^{n} |\lambda_i - \lambda_k|^2} \frac{\prod_{k=1,k\neq i}^{n} |\gamma_j - \lambda_k|^2}{\prod_{k=1,k\neq j}^{n} |\gamma_j - \gamma_k|^2}. \tag{10}$$

*Proof.* Define the Cauchy matrix $C = [c_{ij}]_{n\times n}$ with $c_{ij} = \frac{1}{\gamma_i - \lambda_j}$. Then it follows from the inversion formula for Cauchy matrices in [6] that $C^{-1} = -W\,C^T\,H$, where $W := \operatorname{diag}(w_1, , \dots, w_n)$, $H := \operatorname{diag}(h_1, \dots, h_n)$, and

$$w_i := \frac{\prod_{k=1}^{n} (\lambda_i - \gamma_k)}{\prod_{k=1,k\neq i}^{n} (\lambda_i - \lambda_k)}, h_i := \frac{\prod_{k=1}^{n} (\gamma_i - \lambda_k)}{\prod_{k=1,k\neq i}^{n} (\gamma_i - \gamma_k)}, \qquad i = 1, \dots, n. \tag{11}$$

Applying formulas (5)–(8) to these special data $A$, $b$ and $\mathcal{P}$ we get $f^T = e^T C^{-1}$. Using the formula for $C^{-1}$ we get $f_i = h_i$, $k = 1, \dots, n$, where $f_i$ is the $i$th component of $f$. With the formulas of $h_i$ in (11) we obtain (9).

Using the definition of $\kappa$ in (8) and noticing that $C$ is an eigenvector matrix of $A - bf^T = \Gamma - ef^T$, we only need to normalize the columns of $C$ to one. Let $U := \operatorname{diag}(\mu_1, \dots, \mu_n)$, where

$$\mu_i := \sqrt{\sum_{k=1}^{n} \frac{1}{|\gamma_k - \lambda_i|^2}} = \sqrt{\frac{\sum_{l=1}^{n} \prod_{k=1,k\neq l}^{n} |\gamma_k - \lambda_i|^2}{\prod_{k=1}^{n} |\gamma_k - \lambda_i|^2}},$$

and let $C_0 := CU^{-1}$; then we get $\kappa_F^2 = \|C_0\|_F^2 \, \|C_0{}^{-1}\|_F^2$. Clearly $\|C_0\|_F^2 = n$, and using the formulas for $C^{-1}$ and $U$ we get (10). □

*Remark* 1. Let $\psi(t) = \prod_{k=1}^{n}(t - \lambda_k)$ be the characteristic polynomial of $\Gamma - ef^T$. It is easy to check with $f_i = h_i$ as in (11) that the components of $f$ are just the coefficients of the Lagrange interpolating polynomial for the $n$ points $\{\gamma_k, \psi(\gamma_k)\}_{k=1}^{n}$, i.e., the polynomial $\eta(t) = \sum_{k=1}^{n} f_k \prod_{l=1,l\neq k}^{n}(t-\gamma_l)$ satisfies $\eta(\gamma_k) = \psi(\lambda_k)$. Moreover we have $\psi(t)-\eta(t) = \prod_{k=1}^{n}(t-\gamma_k) =: \phi(t)$, which is just the characteristic polynomial of $\Gamma$.

In Theorem 2.1 we have obtained $\kappa_F$ and $\|f\|_2$ in terms of the pole set and the spectrum of $\Gamma$. The evaluation of the polynomial $\|f\|_2^2$ and the rational function $\kappa_F^2$ in an optimization code is relatively simple; however, there are still difficulties when $n$, the size of the problem, is large. The second difficulty in employing an optimization procedure is the selection of the initial value. A bad initial value will lead to an extremely large $\mathcal{S}$ and for large $n$ this may lead to overflow in the computations. So even if $\mathcal{P}$ exists such that the given SIPP problem is well conditioned, it will be difficult to start the optimization procedure. The third difficulty is that for some systems $(A, b)$ and sets $\mathcal{D}$, even the OSIPP problems is ill conditioned, as we will show in Example 5. In such a case there is no need to use an optimization procedure. Theorem 2.1 also shows that the minimum of $\mathcal{S}$ approaches infinity if some $|\lambda_i| \to \infty$, since in this case $\|f\| \to \infty$ and $\kappa \geq 1$, so $\mathcal{S} \to \infty$.

The following result shows that if $A$ has all distinct and simple eigenvalues, then it is usually sufficient to restrict our discussion to the special case $(\Gamma, e, \mathcal{P})$.

THEOREM 2.2. *Let $(A, b)$ be controllable, $A = X\Gamma X^{-1}$, $\Gamma := \operatorname{diag}(\gamma_1, \ldots, \gamma_n)$, and let $\mathcal{P} = \{\lambda_1, \ldots, \lambda_n\}$ have distinct elements. Denote by $f_d$, $f$ the unique feedback gains of $(A, b, \mathcal{P})$ and $(\Gamma, e, \mathcal{P})$, respectively, and by $\kappa_F^d$, $\kappa_F$ denote the associated scaled spectral condition numbers of $A - bf_d^T$ and $\Gamma - ef^T$ in Frobenius norm. Let $\tilde{b} = \begin{bmatrix} \tilde{b}_1 & \cdots & \tilde{b}_n \end{bmatrix}^T := X^{-1}b$ and $B := \operatorname{diag}(\tilde{b}_1, \ldots, \tilde{b}_n)$. Then*

$$\frac{\|f\|_2}{\|XB\|_2} \leq \|f_d\|_2 \leq \|(XB)^{-1}\|_2\|f\|_2 \tag{12}$$

*and*

$$\frac{\kappa_F}{\sqrt{n}\,\|XB\|_2\,\|(XB)^{-1}\|_2} \leq \kappa_F^d \leq \sqrt{n}\,\|XB\|_2\,\|(XB)^{-1}\|_2\kappa_F. \tag{13}$$

*Proof.* Let

$$A - bf_d^T = G\Lambda G^{-1}, \quad \Lambda = \operatorname{diag}(\lambda_1, \ldots, \lambda_n), \quad \kappa_F^d = \|G\|_F\,\|G^{-1}\|_F,$$

i.e., the columns of $G$ have unit norm. With $A = X\Gamma X^{-1}$ and $B$ defined as above, we have

$$\Gamma - Bef_d^T X = X^{-1}G\Lambda G^{-1}X.$$

Since $(A, b)$ is controllable if and only if $(\Gamma, e)$ is, we obtain (see also [12]) $\tilde{b}_i \neq 0$, $i = 1, \ldots, n$, so $B$ is nonsingular. Performing a similarity transformation with $B^{-1}$, $B$, and using that $B$ and $\Gamma$ are diagonal, we have

$$\Gamma - ef_d^T XB = (XB)^{-1}G\Lambda((XB)^{-1}G)^{-1}.$$

By the uniqueness of the feedback gain, we then have $f^T = f_d^T XB$, which implies (12). Let $C_0$ be defined as in the proof of Theorem 2.1; then $C_0 = (XB)^{-1}GZ$, for some diagonal matrix $Z = \operatorname{diag}(z_1, \ldots, z_n)$ so that the columns of $C_0$ have unit norm. So we have

$$\sqrt{n} = \|C_0\|_F \ge \frac{\|GZ\|_F}{\|XB\|_2} = \frac{\sqrt{\sum_{k=1}^n |z_k|^2}}{\|XB\|_2} \ge \frac{\max_k |z_k|}{\|XB\|_2} = \frac{\|Z\|_2}{\|XB\|_2}, \tag{14}$$

i.e., $\|Z\|_2 \le \sqrt{n}\,\|XB\|_2$. Similarly, by using $G = XBC_0Z^{-1}$, we get $\|Z^{-1}\|_2 \le \sqrt{n}\|(XB)^{-1}\|_2$. Since $G^{-1} = ZC_0{}^{-1}(XB)^{-1}$, we obtain from (14) that

$$\frac{\|C_0^{-1}\|_F}{\sqrt{n}\|(XB)^{-1}\|_2\,\|XB\|_2} \le \|G^{-1}\|_F \le \sqrt{n}\,\|XB\|_2\,\|(XB)^{-1}\|_2\|C_0^{-1}\|_F.$$

Using $\kappa_F^d = \sqrt{n}\|G^{-1}\|_F$, $\kappa_F = \sqrt{n}\|C_0{}^{-1}\|_F$, we obtain (13). □

We see from this result that since $\|XB\|\,\|(XB)^{-1}\|$ is independent of the chosen poles, we can restrict ourselves to the special case $(\Gamma, e, \mathcal{P})$. It is obvious that (12) and (13) and also all subsequent bounds can be extended to the case when $(A, b)$ is controllable and $A$ is diagonalizable.

To analyze the problem further in this case, we will give several bounds for $\mathcal{S}$, as defined in (4).

Theorem 2.3. *Let $\Gamma := \operatorname{diag}(\gamma_1, \ldots, \gamma_n)$, let $(\Gamma, e)$ be controllable, and let $\mathcal{P} = \{\lambda_1, \ldots, \lambda_n\}$ with distinct elements. Suppose that $\lambda(\Gamma) \cap \mathcal{P} = \emptyset$ and set $d_u = \max_{i,j} |\gamma_i - \lambda_j|$, $d_l = \min_{i,j} |\gamma_i - \lambda_j|$. Furthermore, set $w := [w_1, \ldots, w_n]^T$ with $w_i = \frac{\prod_{k=1}^n (\lambda_i - \gamma_k)}{\prod_{k=1,k\neq i}^n (\lambda_i - \lambda_k)}$. Then*

$$n\,\frac{\|w\|_2\,\|f\|_2\sqrt{1 + \|f\|_2^2}}{d_u^2} \le \mathcal{S}_F \le n\,\frac{\|w\|_2\,\|f\|_2\sqrt{1 + \|f\|_2^2}}{d_l^2}. \tag{15}$$

*Proof.* Considering the formulas for $\kappa_F$ and $f$ in Theorem 2.1 and (11), we obtain

$$\begin{aligned}
\kappa_F^2 &= n \sum_{i,j=1}^n \frac{\sum_{l=1}^n \prod_{k=1,k\neq l}^n |\lambda_i - \gamma_k|^2}{\prod_{k=1,k\neq i}^n |\lambda_i - \lambda_k|^2} \frac{\prod_{k=1,k\neq i}^n |\gamma_j - \lambda_k|^2}{\prod_{k=1,k\neq j}^n |\gamma_j - \gamma_k|^2} \\
&= n \sum_{i,j=1}^n \frac{\prod_{k=1}^n |\lambda_i - \gamma_k|^2}{\prod_{k=1,k\neq i}^n |\lambda_i - \lambda_k|^2} \left( \sum_{l=1}^n \frac{1}{|\lambda_i - \gamma_l|^2} \right) \frac{1}{|\gamma_j - \lambda_i|^2} \frac{\prod_{k=1}^n |\gamma_j - \lambda_k|^2}{\prod_{k=1,k\neq j}^n |\gamma_j - \gamma_k|^2} \\
&= n \sum_{i,j=1}^n \left( \sum_{l=1}^n \frac{1}{|\lambda_i - \gamma_l|^2} \right) \frac{1}{|\gamma_j - \lambda_i|^2} |w_i|^2 |f_j|^2.
\end{aligned}$$

Since for all $i$, $j$ we have $d_l \le |\lambda_i - \gamma_j| \le d_u$, it follows that

$$\frac{n}{d_u^4} \le \left( \sum_{l=1}^n \frac{1}{|\lambda_i - \gamma_l|^2} \right) \frac{1}{|\gamma_j - \lambda_i|^2} \le \frac{n}{d_l^4},$$

and hence

$$\frac{n^2}{d_u^4} \sum_{i,j=1}^n |w_i|^2 |f_j|^2 \le \kappa_F^2 \le \frac{n^2}{d_l^4} \sum_{i,j=1}^n |w_i|^2 |f_j|^2.$$

Thus,

$$\frac{n}{d_u^2}\left\|w\right\|_2\left\|f\right\|_2 \le \kappa_F \le \frac{n}{d_l^2}\left\|w\right\|_2\left\|f\right\|_2,$$

and multiplying with $\sqrt{1+\|f\|_2^2}$ yields the conclusion. □

The quantities $d_l$, $d_u$ are the smallest and largest distances between the sets $\Lambda(\Gamma)$ and $\mathcal{P}$. If $d_l << d_u$, in particular when $d_l$ is very small, the upper bound in (15) will usually be an overestimate. But if $d_u/d_l$ is not too large, (15) will be a good estimate for $\mathcal{S}$.

Note that $w$ is the feedback gain for a SIPP problem with $A = \operatorname{diag}(\lambda_1,\ldots,\lambda_n)$, $b = e$, and the pole set $\mathcal{P} = \{\gamma_1,\ldots,\gamma_n\}$. So it has a similar interpretation as $f$ in Remark 1 but in general there is no explicit relationship between $\|w\|$ and $\|f\|$. However, if $\mathcal{P}$ is selected in a particular way, we can get $\|w\| = \|f\|$.

COROLLARY 2.4. *Let $\Gamma := \operatorname{diag}(\gamma_1,\ldots,\gamma_n)$ and let $(\Gamma, e)$ be controllable. If $\mathcal{P} = \{-\overline{\gamma}_1,\ldots,-\overline{\gamma}_n\}$, $\{\overline{\gamma}_1,\ldots,\overline{\gamma}_n\}$, or $\{-\gamma_1,\ldots,-\gamma_n\}$, then $\|w\|_2 = \|f\|_2$ and*

$$\frac{n}{d_u^2}\left\|f\right\|_2^2\sqrt{1+\|f\|_2^2} \le \mathcal{S}_F \le \frac{n}{d_l^2}\left\|f\right\|_2^2\sqrt{1+\|f\|_2^2}. \tag{16}$$

*Proof.* We consider just the case when $\mathcal{P} = \{-\overline{\gamma}_1,\ldots,\overline{\gamma}_n\}$; the other two cases are analogous.

Now $\lambda_i = -\overline{\gamma}_i$, $i = 1,\ldots,n$, so

$$\begin{aligned} w_i &= \frac{\prod_{k=1}^n(\lambda_i-\gamma_k)}{\prod_{k=1,k\neq i}^n(\lambda_i-\lambda_k)} = \frac{\prod_{k=1}^n(-\overline{\gamma}_i-\gamma_k)}{\prod_{k=1,k\neq i}^n(-\overline{\gamma}_i+\overline{\gamma}_k)} \\ &= -\overline{\left(\frac{\prod_{k=1}^n(\gamma_i+\overline{\gamma}_k)}{\prod_{k=1,k\neq i}^n(\gamma_i-\gamma_k)}\right)} = -\overline{\left(\frac{\prod_{k=1}^n(\gamma_i-\lambda_k)}{\prod_{k=1,k\neq i}^n(\gamma_i-\gamma_k)}\right)} = -\overline{f}_i. \end{aligned}$$

Hence $\|w\|_2 = \|f\|_2$, and then (16) follows from (15). □

Since the solution of the SIPP problem consists of the computation of $f$, it is natural that we try to estimate $\kappa$ in terms of $\|f\|$.

Such a result is useless for an optimization procedure since the poles are fixed, but it is valuable in some typical cases arising in applications, i.e., the cases when $\mathcal{P}$ is chosen such that the eigenvalues of $A$ are reflected at the real axis, imaginary axis, or origin, as for example in the well-known Lyapunov method; see, e.g., [9]. A (lower) bound for $\mathcal{S}$ in terms of $\|f\|$ is the following.

THEOREM 2.5. *Consider the SIPP problem with data $A$, $b$, $\mathcal{P}$, where $(A,b)$ is controllable and the poles in $\mathcal{P}$ are distinct. Then*

$$\mathcal{S}_2 \ge \frac{\|b\|_2}{\sqrt{\sum_{i=1}^n\|A-\lambda_i I\|_2^2}}\left\|f\right\|_2\sqrt{1+\|f\|_2^2}. \tag{17}$$

*Proof.* In this case we can apply formulas (5)–(8). From $f^T = \alpha^T G^{-1}$ we get $\|f\|_2 \le \|\alpha\|_2\left\|G^{-1}\right\|_2$, i.e. ,

$$\|G^{-1}\|_2 \ge \frac{\|f\|_2}{\|\alpha\|_2}.$$

Now $\alpha_i$, the $i$th component of $\alpha$ together with $u_i$, the $i$th column of $G$, with $\|u_i\|_2 = 1$, satisfies

$$[A-\lambda_i I, b]\begin{bmatrix} u_i \\ -\alpha_i \end{bmatrix} = 0, \quad \forall i = 1,\ldots,n.$$

Thus, it follows that

$$\|b\|_2^2\,|\alpha_i|^2 = \|(A-\lambda_i I)u_i\|_2^2 \le \|A-\lambda_i I\|_2^2\,\|u_i\|_2^2 = \|A-\lambda_i I\|_2^2\,,$$

which means that $|\alpha_i| \le \frac{\|A-\lambda_i I\|_2}{\|b\|_2}$. So

$$\|\alpha\|_2 \ \le \frac{\sqrt{\sum_{i=1}^n \|A-\lambda_i I\|_2^2}}{\|b\|_2},$$

and by using $\|G\|_2 \ge \max_i \|u_i\|_2 = 1$, we finally get

$$\kappa_2 \ge \|G\|_2\,\|G^{-1}\|_2 \ge \|G\|_2\,\frac{\|f\|_2}{\|\alpha\|_2} \ge \frac{\|b\|_2\,\|f\|_2}{\sqrt{\sum_{i=1}^n \|A-\lambda_i I\|_2^2}}. \qquad \square$$

This lower bound may be very weak. For example, if all poles are selected in a small neighborhood of a single point $\lambda_i$, then $\|f\|_2$ is bounded but $\kappa$ will be very large. But nevertheless, Theorem 2.5 gives a cheap way to estimate $\mathcal{S} = \kappa\sqrt{1+\|f\|^2}$. If, for example, the computed $\|f\|$ is large, say $\|f\| > 1/\sqrt{eps}$, where *eps* is the machine epsilon, then $\mathcal{S} > c/eps$ for some constant $c$. In this case, we can expect that the computed results have lost all significant digits. Consider again special cases where we have explicit solutions.

THEOREM 2.6. *Let* $\Gamma := \operatorname{diag}(\gamma_1,\dots,\gamma_n)$, *with* $\gamma_j \in \mathcal{C}_0^+$, $j = 1,\dots,n$ *and assume that* $(\Gamma, e)$ *is controllable.*

(a) *If we require that* $\mathcal{P} = \{\lambda_1,\dots,\lambda_n\} \subset \mathcal{C}_{-\rho}^-$, *i.e.,* $\operatorname{Re}\lambda_j \le -\rho$, $j = 1,\dots,n$ *for a given real number* $\rho > 0$, *then* $\min\|f\|_2$ *is obtained when* $\operatorname{Re}\lambda_j = -\rho$, *for all* $j = 1,\dots,n$.

(b) *If the* $\gamma_j$ *are such that* $\operatorname{Re}\gamma_j + \rho \ge |\operatorname{Im}\gamma_j|$, *for* $j = 1,\dots,n$, *and we require that the set of poles is as in* (a) *and closed under conjugation, then*

$$\min\|f\|_2 = \sqrt{\sum_{j=1}^n \frac{|\gamma_j+\rho|^{2n}}{\prod_{k=1,k\ne j}^n |\gamma_j-\gamma_k|^2}}, \tag{18}$$

*and the corresponding optimal poles satisfy* $\lambda_j = -\rho$, *for all* $j = 1,\dots,n$.

*Proof.* (a) Let $\gamma_j = a_j + ib_j$, $\lambda_j = x_j + iy_j$, where $a_j$, $b_j$, $x_j$, $y_j$ are real. Then

$$\|f\|_2^2 = \sum_{j=1}^n \frac{\prod_{k=1}^n |\gamma_j-\lambda_k|^2}{\prod_{k=1,k\ne j}^n |\gamma_j-\gamma_k|^2} = \sum_{j=1}^n \frac{\prod_{k=1}^n ((a_j-x_k)^2+(b_j-y_k)^2)}{\prod_{k=1,k\ne j}^n ((a_j-a_k)^2+(b_j-b_k)^2)}.$$

Since $a_j \ge 0$, $x_j \le -\rho$ for all $j$, a necessary condition for a minimum is that $(a_j-x_k)^2$ is minimal for all $k$, which is clearly the case if $x_k = -\rho$, for all $k = 1,\dots,n$.

(b) From (a) we obtain that at a minimum all poles have real part $-\rho$. Suppose that, at the minimum, there exists a pole with nonzero imaginary part $\lambda_s = -\rho + iy_s$. Since the pole set is closed under conjugation, we obtain, for each $\gamma_j$,

$$\begin{aligned}|\gamma_j-\lambda_s|^2|\gamma_j-\bar\lambda_s|^2 &= ((a_j+\rho)^2+(b_j-y_s)^2)((a_j+\rho)^2+(b_j+y_s)^2)\\ &= y_s^4 + 2((a_j+\rho)^2-b_j^2)y_s^2 + ((a_j+\rho)^2+b_j^2)^2.\end{aligned}$$

By assumption, $a_j + \rho \ge |b_j|$, and thus we have

$$\min_{y_s} |\gamma_j-\lambda_s|^2|\gamma_j-\bar\lambda_s|^2 = ((a_j+\rho)^2+b_j^2)^2,$$

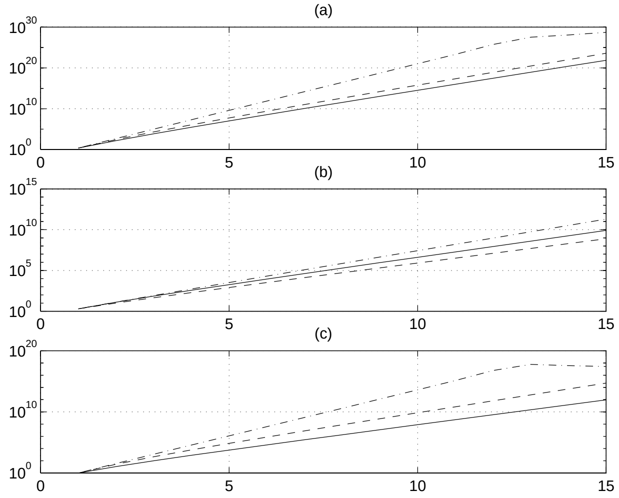

FIG. 4. *Conditioning in Example* 5: (a) $\mathcal{S}$*;* (b) $\|f\|$*;* (c) $\kappa$*; optimal:* —, $\mathcal{P}_1$ : - - -, $\mathcal{P}_2$ : — · —·

i.e., the minimum occurs for $y_s = 0$.

Since each component $|f_k|$ must include a factor of the form $|\gamma_k - \lambda_s||\gamma_k - \bar{\lambda}_s|$, to minimize $\|f\|_2$ we must have $y_s = 0$, which is a contradiction to our assumption. Consequently we have $\lambda_j = -\rho$ for all $j$ and hence we obtain (18). □

Part (b) of Theorem 2.6 shows that, in a particular situation, to get a minimal $\|f\|$ is equivalent to getting the worst conditioning for the closed-loop matrix (in the sense of eigenvalue perturbation theory) since we have to place all poles in one point and the controllability forces the closed-loop matrix to be similar to an $n \times n$ Jordan block. This result also explains the extreme ill conditioning of Example 1.

We also see that, in the situation of Theorem 2.6 (a), at least half of the variables can be removed and that the minimization problem for $\|f\|$ is restricted to a line rather than a half plane. In this situation, suppose that $\mathcal{P}_0 \subset \mathcal{C}^-_{-\rho}$ is a pole set that minimizes $\mathcal{S}$ and let $f_0$ be the feedback gain obtained with $\mathcal{P}_0$. Then with Theorems 2.5 and 2.6 we obtain

$$(19) \qquad \min_{\mathcal{P} \subset \mathcal{C}^-_{-\rho}} \mathcal{S} \geq c \|f_0\|^2 \geq c \min_{\mathcal{P} \subset \{-\rho + yi | y \in \mathcal{R}\}} \|f\|^2 ,$$

for a constant $c$ determined by $A$, $b$, $\mathcal{P}_0$. Thus, if $\min \|f\|$ is large, then the OSIPP problem is incurably ill conditioned and we cannot hope to improve the problem by choosing the poles. Consider the following example.

*Example* 5. Let $A = \text{diag}(1, \ldots, n)$, $b = e$, $\mathcal{P} \subset \mathcal{C}^-_{-1}$, and let $n$ vary from 1 to 15. We used a heuristic "random search" algorithm to choose the set $\mathcal{P}_0$ that minimizes $\mathcal{S}$ and determine a minimal value for $\mathcal{S}$. The resulting condition numbers are shown in Figure 4, as well as the related numbers $\kappa$ and $\|f\|$. For comparison we also display the three numbers $\|f\|$, $\kappa$, and $\mathcal{S}$ for the pole sets $\mathcal{P}_1 = \{-1 - \frac{n-1}{2}i, -1 - \frac{n-3}{2}i, \ldots, -1 + \frac{n-3}{2}i, -1 + \frac{n-1}{2}i\}$ and $\mathcal{P}_2 = \{-1, -2, \ldots, -n\}$, respectively.

We see that the magnitude of $\mathcal{S}$ can be reduced, but even so, when $n = 11$, $\mathcal{S}_{opt} = 10^{16}$. For $n = 15$, $\mathcal{S}_{opt} = 7.5 \times 10^{21}$ and 10 eigenvalues of $A - bf^T$ are in $\mathcal{C}^+_0$. So for $n \geq 15$, it is impossible to place the poles to the left of the line $-1 + yi$ via a numerical procedure. We can also check the ill conditioning by the results in Theorems 2.5 and 2.6. Actually, for $n = 15$, $\min \|f\| = 3.45 \times 10^8$.

Table 3
*Eigenvalue error and eps $\cdot \mathcal{S}$.*

| $n$ | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $eps \cdot \mathcal{S}$ | 5.0e-16 | 3.3e-14 | 1.6e-12 | 6.5e-11 | 2.3e-9 | 7.9e-8 | 2.5e-6 | 8.0e-5 | 0.0024 | 0.0741 | 2.2217 | 66.0347 | 2.0e+3 | 5.7e+4 | 1.7e+6 |
| $E_n$ | 0 | 1.8e-15 | 6.7e-14 | 1.8e-12 | 6.7e-11 | 1.2e-10 | 3.7e-9 | 3.6e-8 | 1.1e-7 | 3.4e-5 | 3.5e-4 | 4.2e-3 | 6.0e-2 | 0.6371 | 7.132 |

To illustrate again the importance of $\mathcal{S}$, in Table 3 we list $eps{\cdot}\mathcal{S}$ and the eigenvalue errors for the poles obtained from our heuristic search algorithm for different $n$. Here $E_n = \max_j |\mu_j - \lambda_j|$, where $\Lambda(A - b\tilde{f}^T) = \{\mu_1, \ldots, \mu_n\}$, and $\tilde{f}$ is the computed feedback with these poles. We see that the error in the eigenvalues grows roughly in the same way as $eps \cdot \mathcal{S}$. Observe that $\|f\|$ is smaller in case $\mathcal{P}_1$ than in the optimal case. As we discussed above, this shows that just minimizing $\|f\|$ is not sufficient to minimize $\mathcal{S}$.

Note that the choice of poles in Theorem 2.6 is quite common. In practice, we often require $\mathcal{P}$ to be in the left half plane so that the closed-loop matrix $A - bf^T$ is *stable*. Also, one often does the pole placement only on the eigenvalues of $A$ with nonnegative real parts, while keeping the rest of the eigenvalues fixed, since this can save a lot of computational work; see [8].

We have seen under the assumptions of Theorem 2.6 that min $\|f\|$ is achieved with poles on the line $-\rho + yi$. From (15) of Theorem 2.3, $\mathcal{S}_F \approx \|w\|_2 \, \|f\|_2^2$. When the pole set $\mathcal{P}$ moves away from the line $-\rho + yi$, $\|f\|$ will increase. Clearly it is possible that $\|w\|$ decreases, but $\|f\|$ enters quadratically in $\mathcal{S}$, so an increasing magnitude of $\|f\|$ usually quickly compensates a decreasing magnitude of $\|w\|$.

Although we are not able to minimize the condition number of the SIPP problem directly, our analysis of the governing factors in this condition number leads us to the following conjecture.

Conjecture. *Suppose that $(A, b)$ is controllable and that $A$ has all eigenvalues in the right half plane. If we require that $\mathcal{P} = \{\lambda_1, \ldots, \lambda_n\} \subset \mathcal{C}^-_{-\rho}$, then the pole placement problem with minimal condition number is achieved when all elements of $\mathcal{P}$ are on the line $-\rho + yi$.*

Such a selection of poles may help to check the conditioning of the SIPP problem and may also be a good choice for the initial poles in an optimization of $\mathcal{S}$.

**3. Numerical experiments.** In this section we will give some further numerical examples which illustrate the theoretical results from the previous sections, and we will also illustrate other factors that contribute to the conditioning and that may be used to improve the bounds. A factors that contributes to the conditioning is the width in the set of imaginary parts of $\Lambda(A)$, but in the case of a stabilization problem, the distance between the real parts of the unstable eigenvalues of $A$ and the desired new locations for these is also a contributing factor.

The observations made in this section are still based only on numerical experiments, but they indicate directions of research.

In all examples, the spectral condition number of the closed loop matrices $\kappa$ were computed by first forming the matrix $G$ in (5) and then applying the Matlab *cond* function. The feedback gain $f$ was generated either via the Miminis–Paige algorithm [14] or via the formulas (5)–(7).

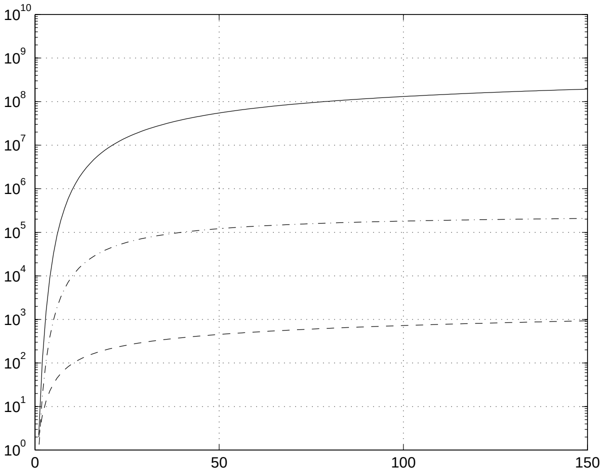

FIG. 5. *Conditioning in Example* 6: *poles reflected at the imaginary axis* $\mathcal{S}$:—, $\|f\|$: - - -, $\kappa$: $-\cdot-$.

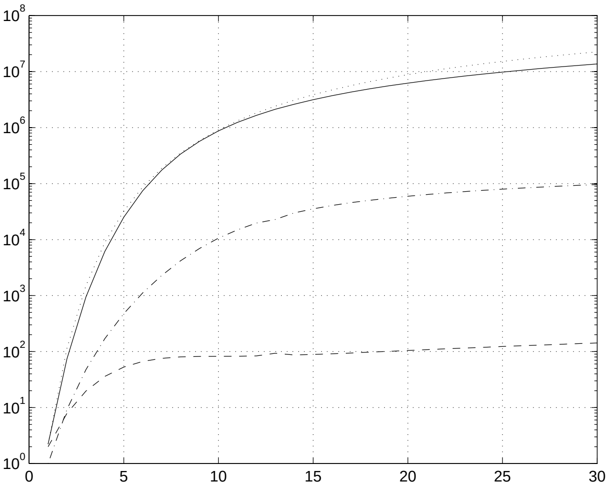

FIG. 6. *Optimal conditioning in Example* 6. $\mathcal{S}$:—, $\mathcal{S}_{ref}$: $\cdots$, $\|f\|$: - - -, $\kappa$: $-\cdot-$.

Our first example demonstrates that a certain geometric relationship between the eigenvalues of $A$ and the chosen poles can lead to a reasonable conditioning.

*Example* 6. Let $A = I_n + i\,\mathrm{diag}(-\frac{n-1}{2}, -\frac{n-3}{2}, \ldots, \frac{n-3}{2}, \frac{n-1}{2})$, $b = e$, and $\mathcal{P} \subset \mathcal{C}^-_{-1}$.

Let $\mathcal{P}$ be the set of poles obtained by reflecting the eigenvalues of $A$ about the imaginary axis, i.e., $\mathcal{P} = \{\lambda_1, \ldots, \lambda_n\}$ and $\lambda_j = -1 + i\frac{n+1-2j}{2}$. The values of $\|f\|$, $\kappa$, and $\mathcal{S}$ with $n = 1 : 150$ are displayed in Figure 5. For comparison we again used a "random search" method for $n = 1 : 30$ to determine a set of "optimal" poles $\mathcal{P}_0$. The condition estimates for $\mathcal{P}_0$ and $\mathcal{P}$ are given in Figure 6. In both cases, $\mathcal{S}$ increases monotonially with $n$, but as $n > 10$, it grows very slowly. In the first case, when $n = 150$, $\mathcal{S} = 1.9362 \times 10^8$.

The magnitude of $\mathcal{S}$ is reduced with the pole set $\mathcal{P}_0$, but just marginally. We also depict $eps \cdot \mathcal{S}$, and the error for the closed-loop eigenvalues $E_n$, in Figures 7 and 8 for both cases.

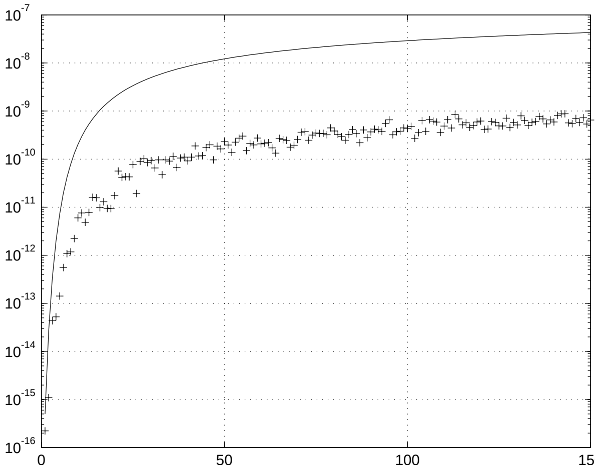

FIG. 7. *Eigenvalue error and* $eps \cdot \mathcal{S}$ *in Example* 6 *with poles reflected at the imaginary axis.* $eig_e : +++$, $eps \cdot \mathcal{S}$ : —.

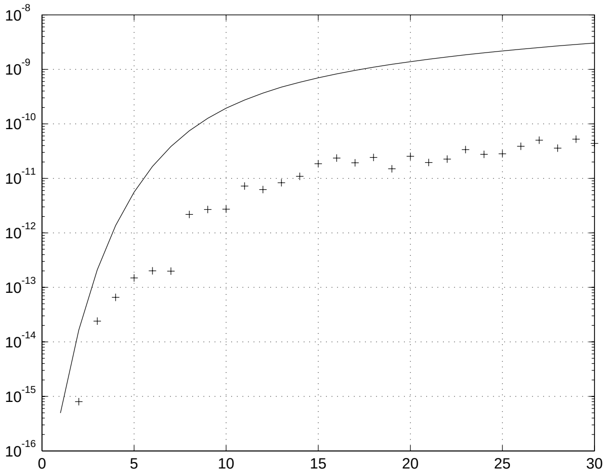

FIG. 8. *Eigenvalue error and* $eps \cdot \mathcal{S}$ *in Example* 6 *with "optimal poles"* $eig_e : +++$, $eps \cdot \mathcal{S}$ : —.

Unlike in Example 5, the conditioning of the SIPP problems in Example 6 seems to be reasonable. This asks for an explanation. Let $A = \rho_1 I_n + i \cdot d \cdot \mathrm{diag}(-\frac{n-1}{2}, \ldots, \frac{n-1}{2})$, $b = e$, and $\mathcal{P} = \{-\rho_2 - d\frac{n-1}{2}i, \ldots, -\rho_2 + d\frac{n-1}{2}i\}$. Suppose that $d > 0$, $\rho_1, \rho_2 > 0$. Introducing $\rho := \rho_1 + \rho_2$ we obtain, for each component of $f$,

$$\begin{aligned}
|f_j|^2 &= \frac{\prod_{k=1}^{n} |\gamma_j - \lambda_k|^2}{\prod_{k=1,k\neq j}^{n} |\gamma_j - \gamma_k|^2} \\
&= \frac{\prod_{k=1}^{n} (\rho^2 + d^2(j - \frac{n+1}{2} - (k - \frac{n+1}{2}))^2)}{d^2 \prod_{k=1,k\neq j}^{n} (j - \frac{n+1}{2} - (k - \frac{n+1}{2}))^2} \\
&= \rho^2 \prod_{k=1,k\neq j}^{n} \left(1 + \frac{(\rho/d)^2}{(j-k)^2}\right)
\end{aligned}$$

$$
\begin{aligned}
&= \rho^2 \prod_{k=1}^{j-1}\left(1+\left(\frac{\rho}{d}\right)^2\frac{1}{k^2}\right)\prod_{k=1}^{n-j}\left(1+\left(\frac{\rho}{d}\right)^2\frac{1}{k^2}\right)\\
&\le \rho^2 e^{(\frac{\rho}{d})^2\sum_{k=1}^{n-j}\frac{1}{k^2}} e^{(\frac{\rho}{d})^2\sum_{k=1}^{j-1}\frac{1}{k^2}} \le \rho^2 e^{4(\frac{\rho}{d})^2}.
\end{aligned}
$$

So, $\|f\|_2 \le \sqrt{n}\rho e^{2(\frac{\rho}{d})^2}$. Similarly, we get $\|w\|_2 \le \sqrt{n}\rho e^{2(\frac{\rho}{d})^2}$ and hence

$$
\mathcal{S}_F \approx \|f\|_2\,\kappa_F \le \frac{n}{\rho}\,\|w\|_2\,\|f\|_2^2 \le n^{\frac{5}{2}}\rho^2 e^{6(\frac{\rho}{d})^2}. \tag{20}
$$

In Example 6 we have $\rho = 2$, $d = 1$, so $\mathcal{S} \le 1.06 n^{\frac{5}{2}} \times 10^{11}$. Clearly this is a large overestimate, but the bound increases polynomially in $n$. Consider a generalized problem as Example 5, with $A = diag(\rho_1 + d, \dots, \rho_1 + (n-1)d)$, $b = e$, and $\mathcal{P} = \{-\rho_2 - d, \dots, -\rho_2 - (n-1)d\}$. Suppose also that $d, \rho_1, \rho_2 > 0$ and let $\rho := \rho_1 + \rho_2$. Then for the $n$th components of $f$ and $w$, we have

$$
\begin{aligned}
|f_n| = |w_n| &= (\rho + 2(n-1)d)\frac{\prod_{k=1}^{n-1}(\rho/d + (n-2+k))}{(n-1)!}\\
&> (\rho + 2(n-1)d)\frac{(2n-3)!}{(n-1)!(n-2)!}.
\end{aligned}
$$

Hence by (15),

$$
\mathcal{S}_F \ge \frac{n}{(\rho + 2(n-1)d)^2}|f_n|^2|w_n| \sim O(4^{3n}).
$$

So $\mathcal{S}_F$ increases exponentially in $n$ regardless of the magnitude of $\frac{\rho}{d}$.

We also see the importance of the scalar $\frac{\rho}{d}$ for the problems in Example 6. The smaller it is, the better conditioned the related SIPP problem will be. Since $\frac{\rho}{d}$ is determined not only by $d$, which describes the width in the set of imaginary parts of the eigenvalues of $A$, but also by the distance between the real parts of $\Lambda(A)$ and those in $\mathcal{P}$, one should observe that $\mathcal{S}$ changes when the real parts of $\Lambda(A)$ vary. Consider the following example.

*Example* 7. Let

$$
\begin{aligned}
A = \frac{1}{2}I_{31} - i\cos\left(\frac{(m-1)\pi}{180}\right)&\operatorname{diag}(-15, -14, \dots, 14, 15)\\
+ \sin\left(\frac{(m-1)\pi}{180}\right)&\operatorname{diag}(15, 14, \dots, 14, 15),
\end{aligned}
$$

$m = 1 : 90$, $b = e$. $\mathcal{P}$ is selected in the following two different ways.

*Case* I. $\mathcal{P} = -\lambda(\bar{A})$.

*Case* II. $\mathcal{P} = \{-0.5 - 15\cos(\frac{(m-1)\pi}{180})i, \dots, -0.5 + 15\cos(\frac{(m-1)\pi}{180})i\}$.

The condition estimates are depicted in Figure 9. As $m$ increases, $\rho$ grows and $d$ decreases, so $\mathcal{S}$ grows, too. The test results support this observation.

We have so far mostly considered matrices $A$ with regular eigenvalue patterns, but the same behavior is observed in the general case.

From all examples that we have tested, we see that the wider the set of imaginary parts of $\Lambda(A)$ and the smaller the distance between the set of real parts of $\Lambda(A)$ to that of $\mathcal{P}$, the better conditioning of the corresponding SIPP problem we can expect.

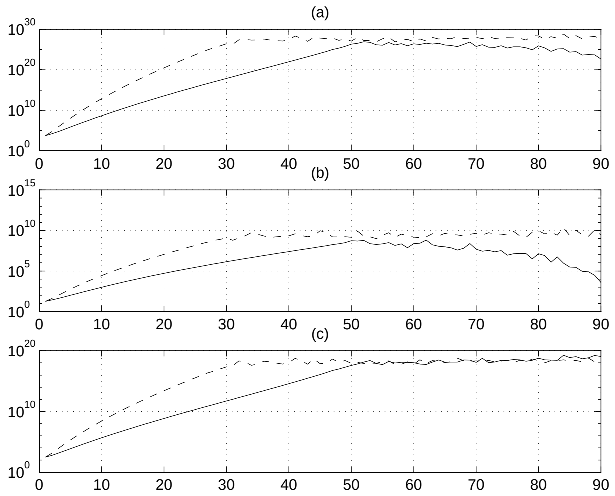

FIG. 9. *Conditioning in Example* 7: (a) $\mathcal{S}$; (b) $\|f\|$; (c) $\kappa$; *Case* I: - - -, *Case* II: —.

With these observations, the class of SIPP problems in Example 5 is probably the worst class of problems, while the class of problems in Example 6 is probably the best class.

In many control problems, for example in robust stabilization, the choice of the imaginary parts of $\Lambda(A)$ seems to be unimportant. But the above observations indicate that the imaginary parts are of great importance since they participate heavily in the conditioning of the SIPP problem.

We also tested many examples with multiple poles in $\mathcal{P}$, but then, as could be expected, the conditioning is much worse than in the case of distinct poles. If the matrix $A$ has eigenvalues with negative real parts, which can also be moved to improve the conditioning, then it can be predicted that $\mathcal{S}$ can be reduced compared to the case when just the eigenvalues with nonnegative real parts are assigned. Numerical examples support this prediction. For more examples, see [13].

## 4. Conclusions.

In this paper we have studied the problem of optimizing the conditioning of the single input pole placement problem when the poles are allowed to vary in specific regions of the complex plane. It is in general very difficult to minimize the condition number or the bounds for the eigenvalue error (or the error in $f$) in the SIPP problem, since these functions are very complicated and, in particular, for large $n$ a numerical minimization seems prohibitive. To get arround this difficulty, we have studied two of the factors in the perturbation bounds, $f$ and $\kappa$, the scaled spectral condition number of the closed-loop matrix and, we have neglected the effects of the condition number of the matrix $A$ and the distance to uncontrollability of $(A, b)$.

For problems where the optimization of $f$ and $\kappa$ can be carried out explicitly, we have determined formulas for the minima. From these formulas we are motivated to conjecture the location for the optimal pole selection in the important stabilization problem, where $\Lambda(A) \subset \mathcal{C}_0^+$, $\mathcal{P} \subset \mathcal{C}_{-\rho}^-$.

By several numerical tests we have indicated how the conditioning of the SIPP problem is determined by the distribution of the eigenvalues of $A$ and the geometric relationship to the selected poles.

Also, we have pointed out that in order to study the accuracy of the results of the SIPP problem, it is not enough to consider just the accuracy of the feedback gain.

In Example 4, the computed gain vectors have no correct digits, while the associated eigenvalues of the computed closed-loop matrix still have about eight correct digits. But as shown by Example 1, the converse may also be the case, i.e., even though $f$ is very accurate, the poles of the computed closed-loop system are far away from the desired poles.

**Acknowledgment.** We thank an anonymous referee for helpful comments which improved the presentation and readability of the paper.

## REFERENCES

[1] J. Ackermann, *Der Entwurf linearer Regelungssysteme im Zustandsraum*, Regelungstechnik und Prozessdatenverarbeitung, 7 (1972), pp. 297–300.

[2] M. Arnold, *Algorithms and Conditioning for Eigenvalue Assignment*, Ph.D. thesis, Dept. of Mathematics, Northern Illinois University, De Kalb, IL, 1993.

[3] D. Boley, *Estimating the sensitivity of the algebraic structure of pencils with simple eigenvlaue estimates*, SIAM J. Matrix Anal. Appl., 11 (1990), pp. 632–643.

[4] W. L. Brogan, *Applications of a determinant identity to pole-placement and observer problems*, IEEE Trans. Automat. Control, AC-19 (1974), pp. 612–614.

[5] J. Demmel, *The condition number of equivalence transformations that block diagonalize matrix pencils*, SIAM J. Numer. Anal., 20 (1983), pp. 599–610.

[6] T. Finck, G. Heinig, and K. Rost, *An inversion formula and fast algorithms for Cauchy-Vandermonde matrices*, Linear Algebra Appl., 183 (1993), pp. 179–191.

[7] G. H. Golub and C. F. Van Loan, *Matrix Computations*, 2nd ed., Johns Hopkins Univ. Press, Baltimore, MD, 1989.

[8] C. He, A. J. Laub, and V. Mehrmann, *Placing Plenty of Poles is Pretty Preposterous*, preprint 95 − 17, DFG-Forschergruppe Scientific Parallel Computing, Fak. f. Mathematik, TU Chemnitz-Zwickau, D-09107 Chemnitz, FRG, May 1995.

[9] C. He and V. Mehrmann, *Stabilization of large linear systems*, in Proceedings IEEE Workshop on Computer-Intensive Methods in Control and Signal Processing, Prague, September 1994, M. Karny and K. Warwick, eds. IEEE Computer Society Press, Los Alamitos, CA, pp. 91–100; also see preprint 94 − 21, DFG-Forschergruppe Scientific Parallel Computing, Fak. f. Mathematik, TU Chemnitz-Zwickau, D-09107 Chemnitz, FRG, October 1994.

[10] J. Kautsky, N. K. Nichols, and P. Van Dooren, *Robust pole assignment in linear state feedback*, Internat. J. Control, 41 (1985), pp. 1129–1155.

[11] M. M. Konstantinov and P. Hr. Petkov, *Conditioning of Linear State Feedback*, Technical report 93-61, Dept. of Engineering, Leicester University, 1993.

[12] V. Mehrmann and H. Xu, *An analysis of the pole placement problem.* I. *The single-input case*, ETNA, 4 (1996), pp. 89–105.

[13] V. Mehrmann and H. Xu, *Choosing Poles So That The Single-Input Pole Placement Problem Is Well-conditioned*, preprint SFB 393/96-01, Sonderforschungsbereich 393, Numerische Simulation auf massivparallelen Rechnern, Fak. f. Mathematik, TU Chemnitz-Zwickau, D-09107 Chemnitz, FRG, January 1996.

[14] G. S. Miminis and C. C. Paige, *A direct algorithm for pole assignment of time-invariant multi-input linear systems using state feedback*, Automatica, 24 (1988), pp. 343–356.

[15] P. Hr. Petkov, N. D. Christov, and M. M. Konstantinov, *A computational algorithm for pole assignment of linear multiinput systems*, IEEE Trans. Automat. Control, AC-31 (1986), pp. 1044–1047.

[16] V. Sima, *An efficient Schur method to solve the stabilization problem*, IEEE Trans. Automat. Control, AC-26 (1981), pp. 724–725.

[17] J. G. Sun, *Perturbation analysis of the pole assignment problem*, SIAM J. Matrix Anal. Appl., 17 (1996), pp. 313–331.

[18] A. Varga, *A Schur method for pole assignment*, IEEE Trans. Automat. Control, AC-26 (1981), pp. 517–519.

[19] J. H. Wilkinson, *The Algebraic Eigenvalue Problem*, Oxford University Press, Oxford, 1965.

[20] W. Wonham, *Linear Multivariable Control. A Geometric Approach*, Springer-Verlag, Berlin, Heidelberg, 1974.