Interior Point Methods Meet Simplex in L_∞ Fitting Problems

Aurelio R. L. Oliveira * Christiano Lyra †

October 27, 2004

Abstract

Interior point methods specialized to the L_∞ fitting problem are surveyed, improved and compared with the traditional simplex approach. A primal affine-scaling interior point method is presented, completing the affine-scaling interior point family approach to the L_∞ fitting problem. Computational complexity and data storage are reduced for interior point approaches when dealing with polynomial fitting problems. Numerical experiments indicate that interior point approaches rarely perform better than the simplex method for the tested problems. The primal affine-scaling method presented in this paper achieved the best results among the interior point family.

Keywords: L_∞ curve fitting, Chebychev norm-based fitting problem, L_∞ based statistical analysis, interior point methods, linear programming.

1 Introduction

The L_∞-norm based curve fitting method for data analysis minimizes the maximum absolute error. It has the intuitive appeal of protection against

*Applied Mathematics Department (DMA), State University of Campinas (UNICAMP), C.P. 6065, 13083-970 Campinas, SP, Brazil; aurelio@ime.unicamp.br

†School of Electrical Engineering and Computer Science (FEEC), State University of Campinas (UNICAMP), R. Albert Einstein 400, 13083-970 Campinas, SP, Brazil; chrlyra@densis.fee.unicamp.br

"the worst" and is usually recommended for short-tailed data distribution that appears in practice due to unconscious or deliberate truncations (Narula and Wellington, 1988).

In late fifties, Wagner (1959) provided a linear programming representation for the L_∞ fitting problem. Such a representation was motivated by the attractive practical way of solving problems with the simplex algorithm. Teamed with specially designed simplex procedures, it played a major role in spreading the use of L_∞ to statistical analysis. After the interior point ideas appeared in the optimization scene (Karmarkar, 1984), it was a matter of time to have specially tailored interior point methods developed for both L_1 and L_∞ curve fitting problems.

A previous paper revised and enhanced interior point approaches to L_1 problems (Oliveira, Nascimento and Lyra, 2000). Extensive numerical experience backed up the conclusion that this family of methods behaves better than the simplex method for L_1 fitting. This paper revisits several interior point methods for the L_∞ fitting problem proposed in recent years. A primal affine-scaling approach is presented, completing the affine-scaling interior point alternatives to L_∞ fitting problems.

Polynomial L_∞ fitting problems are discussed in detail. Their special structure enables improvements in computational complexity and data storage requirements for all interior point methods.

Benchmarks compare all interior point approaches against one of the best known implementations of specialized simplex based procedures for L_∞ fitting problems (Barrodale and Phillips, 1975). Coleman and Li's (1992) method, closely related to the interior-point family, is also considered.

The paper is organized as follows. The next section presents a linear programming formulation for L_∞ fitting problems. Section 3 discusses the main interior-point approaches and Coleman-Li's method for L_∞ fitting problems. Section 4 presents enhancements to the polynomial fitting problem. Benchmarks are discussed in Section 5. Conclusions follow.

2 The L_∞ Fitting Problem Formulation

Consider the L_∞ fitting problem,

$$\begin{array}{ll} \text{minimize} & \|r\|_\infty \\ \text{subject to} & Ax + r = b \end{array} \tag{1}$$

where A is a full column rank $n \times m$ matrix and r, b and x are column vectors of appropriate dimension.

Wagner (1959) and Arthanari and Dodge (1981) proposed alternative linear programming formulations for problem (1). In this paper, we chose to work with the former, which is the more generally adopted. However, the results presented are valid for either formulation. According to Wagner's formulation, the L_∞ fitting problem is transformed into the following linear programming problem:

$$\begin{array}{ll} \text{minimize} & \beta \\ \text{subject to} & Ax - e\beta + v = b, \\ & Ax + e\beta - u = b, \quad (u, v) \geq 0, \end{array} \tag{2}$$

where β is a scalar, u and v are column n-vectors and e is the n-vector composed exclusively of ones.

The dual linear program of (2) is formulated as follows:

$$\begin{array}{ll} \text{maximize} & b^t(y - z) \\ \text{subject to} & A^t(y - z) = 0, \\ & e^t(y + z) = 1, \quad (y, z) \geq 0, \end{array} \tag{3}$$

where y and z are n-vectors. Optimal solutions for (2) and (3) satisfy primal and dual feasibility and the *complementary slackness* conditions: $UYe = 0; VZe = 0$, where U, V, Y and Z are diagonal matrices with the diagonal entries given by u, v, y and z, respectively.

3 Interior Point Methods and L_∞ Fitting

Ruzinsky and Olsen (1989) presented the first specialization of interior point approaches for solving (2), a dual affine-scaling method. A primal-dual method and a predictor-corrector variant were proposed by Zhang (1993). To the best of our knowledge, the primal affine-scaling is presented for the first time in this paper.

The method proposed by Coleman and Li (1992) has the attractive theoretical property of quadratic convergence and is pretty much in the flavor of interior point methods. Its iterations are similar to the iterations of interior point methods with only two main differences: the linear system matrix structure and the line search procedure, which are much easier to compute for interior point approaches.

Given β^0, x^0 and $(u^0, v^0) > 0$ such that $Ax^0 - e\beta^0 + v^0 = b$ and $Ax^0 + e\beta^0 - u^0 = b$.
For $k = 0, 1, 2, \ldots$, do

$$
\begin{aligned}
D^k &= ((U^k)^{-2} + (V^k)^{-2}) && (1)\\
E^k &= ((U^k)^{-2} - (V^k)^{-2}) && (2)\\
x_1^k &= -(A^t D^k A)^{-1}(A^t E^k e) && (3)\\
\Delta\beta^k &= -(e^t D^k e + e^t E^k A x_1^k)^{-1} && (4)\\
\Delta x^k &= x_1^k \Delta\beta^k && (5)\\
\Delta u^k &= e\Delta\beta^k + A\Delta x^k && (6)\\
\Delta v^k &= e\Delta\beta^k - A\Delta x^k && (7)\\
\alpha^k &= \tau \min_i \left\{ \min \left\{ \frac{-u_i^k}{\Delta u_i^k} \,|\Delta u_i^k < 0 \,;\, \frac{-v_i^k}{\Delta v_i^k} \,|\Delta v_i^k < 0 \right\}\right\} && (8)\\
w^{k+1} &= w^k + \alpha^k \Delta w^k && (9)
\end{aligned}
$$

Until Convergence.

Figure 1: Primal Affine-Scaling Method (P)

We shall sometimes refer to the methods according to the acronyms defined as follows: P for Primal, D for Dual, PD for Primal-Dual, PC for Predictor-Corrector and C&L for Coleman and Li.

Figures 1, 2, 3 and 4 summarize the interior point approaches specialized to L_∞ fitting problems (1). Observe that the primal affine-scaling method has the smallest overall computational effort by iteration. Figure 5 summarizes C&L method.

In Figures 1 to 5 we sometimes use the notation $r = b - Ax$, $w = (x, u, v)$ and the gap $\gamma = u^t y + v^t z$. We also consider that the solution of a linear system like

$$
\begin{pmatrix} A^t DA & A^t Ee \\ e^t EA & e^t De \end{pmatrix} \begin{pmatrix} \Delta x \\ \Delta\beta \end{pmatrix} = \begin{pmatrix} r \\ \rho \end{pmatrix}
$$

is given by $\Delta x = x_0 + x_1 \Delta\beta$, where $x_0 = (A^t DA)^{-1} r$, $x_1 = -(A^t DA)^{-1} A^t Ee$ and $\Delta\beta = \frac{\rho - e^t EAx_0}{e^t De + e^t EAx_1}$. The right hand side, given by r and ρ, are defined by the respective interior point method. Note that for the dual affine-scaling method, the unknowns of the linear system are an estimation of the primal variables x and β, instead of the search directions Δx and $\Delta\beta$.

The parameters τ and ν used in some of the methods are presented in

Given $(y^0, z^0) > 0$ such that $A^t(y^0 - z^0) = 0$ and $e^t(y^0 + z^0) = 1$.
For $k = 0, 1, 2, \ldots$, do

$$\begin{aligned}
D^k &= (Y^k)^2 + (Z^k)^2 && (1)\\
E^k &= (Y^k)^2 - (Z^k)^2 && (2)\\
x_0^k &= (A^t D^k A)^{-1}(A^t D^k b) && (3)\\
x_1^k &= -(A^t D^k A)^{-1}(A^t E^k e) && (4)\\
\beta^k &= \frac{(e^t E^k b - e^t E^k A x_0^k)}{(e^t D^k e + e^t E^k A x_1^k)} && (5)\\
x^k &= x_0^k + \beta^k x_1^k && (6)\\
\Delta y^k &= (Y^k)^2(\beta^k e - r^k) && (7)\\
\Delta z^k &= (Z^k)^2(\beta^k e + r^k) && (8)\\
\alpha^k &= \tau \min_i \left\{ \min \left\{ \frac{-y_i^k}{\Delta y_i^k} \,\middle|\, \Delta y_i^k < 0 \,;\, \frac{-z_i^k}{\Delta z_i^k} \,\middle|\, \Delta z_i^k < 0 \right\} \right\} && (9)\\
y^{k+1} &= y^k + \alpha^k \Delta y^k && (10)\\
z^{k+1} &= z^k + \alpha^k \Delta z^k && (11)
\end{aligned}$$

Until Convergence.

Figure 2: Dual Affine-Scaling Method (D)

Given β^0, x^0 and $(u^0, v^0) > 0$ such that $Ax^0 - e\beta^0 + v^0 = b$, $Ax^0 + e\beta^0 - u^0 = b$ and $(y^0, z^0) > 0$ such that $A^t(y^0 - z^0) = 0$ and $e^t(y^0 + z^0) = 1$.
For $k = 0, 1, 2, \ldots$, do

$$\begin{aligned}
\mu^k &= \frac{\gamma^k}{(2n)^{\frac{3}{2}}} && (1)\\
D^k &= Y^k(U^k)^{-1} + Z^k(V^k)^{-1} && (2)\\
E^k &= Y^k(U^k)^{-1} - Z^k(V^k)^{-1} && (3)\\
x_0^k &= (A^tD^kA)^{-1}A^t((\mu^k((U^k)^{-1} - (V^k)^{-1})e + z^k - y^k) && (4)\\
x_1^k &= -(A^tD^kA)^{-1}(A^tE^ke) && (5)\\
\Delta\beta^k &= \frac{(e^t((\mu^k((U^k)^{-1} + (V^k)^{-1})e - z^k - y^k) - e^tE^kAx_0^k)}{(e^tD^ke + e^tE^kAx_1^k)} && (6)\\
\Delta x^k &= x_0^k + \Delta\beta^k x_1^k && (7)\\
\Delta u^k &= e\Delta\beta^k + A\Delta x^k && (8)\\
\Delta v^k &= e\Delta\beta^k - A\Delta x^k && (9)\\
\Delta y^k &= \mu^k(U^k)^{-1}e - y^k - Y^k(U^k)^{-1}\Delta u^k && (10)\\
\Delta z^k &= \mu^k(V^k)^{-1}e - z^k - Z^k(V^k)^{-1}\Delta v^k && (11)\\
\hat{\alpha}_p^k &= \min_i\left\{\min\left\{\frac{-u_i^k}{\Delta u_i^k}\,\middle|\,\Delta u_i^k < 0\,;\,\frac{-v_i^k}{\Delta v_i^k}\,\middle|\,\Delta v_i^k < 0\right\}\right\} && (12)\\
\hat{\alpha}_d^k &= \min_i\left\{\min\left\{\frac{-y_i^k}{\Delta y_i^k}\,\middle|\,\Delta y_i^k < 0\,;\,\frac{-z_i^k}{\Delta z_i^k}\,\middle|\,\Delta z_i^k < 0\right\}\right\} && (13)\\
(\alpha_p^k, \alpha_d^k) &= \begin{cases} (\hat{\alpha}_p^k, \hat{\alpha}_p^k), \text{ if } \hat{\alpha}_d^k > \hat{\alpha}_p^k \text{ and } y_k^t\Delta u^k + z_k^t\Delta v^k > 0,\\ (\hat{\alpha}_d^k, \hat{\alpha}_d^k), \text{ if } \hat{\alpha}_p^k > \hat{\alpha}_d^k \text{ and } u_k^t\Delta y^k + v_k^t\Delta z^k > 0,\\ (\hat{\alpha}_p^k, \hat{\alpha}_d^k), \text{ otherwise.}\end{cases} && (14)\\
w^{k+1} &= w^k + \tau^k\alpha_p^k\Delta w^k && (15)\\
y^{k+1} &= y^k + \tau^k\alpha_d^k\Delta y^k && (16)\\
z^{k+1} &= z^k + \tau^k\alpha_d^k\Delta z^k && (17)
\end{aligned}$$

Until Convergence.

Figure 3: Primal-Dual Method (PD)

Given β^0, x^0 and $(u^0, v^0) > 0$ such that $Ax^0 - e\beta^0 + v^0 = b$, $Ax^0 + e\beta^0 - u^0 = b$ and $(y^0, z^0) > 0$ such that $A^t(y^0 - z^0) = 0$ and $e^t(y^0 + z^0) = 1$.
For $k = 0, 1, 2, \ldots$, do

$$
\begin{aligned}
\mu^k &= \frac{\gamma^k}{(2n)^{\frac{3}{2}}} && (1)\\
D^k &= Y^k(U^k)^{-1} + Z^k(V^k)^{-1} && (2)\\
E^k &= Y^k(U^k)^{-1} - Z^k(V^k)^{-1} && (3)\\
\tilde{x}_0^k &= (A^tD^kA)^{-1}A^t(z^k - y^k) && (4)\\
x_1^k &= -(A^tD^kA)^{-1}(A^tE^ke) && (5)\\
\Delta\tilde{\beta}^k &= \frac{(-e^t(z^k + y^k) - e^tE^kA\tilde{x}_0^k)}{(e^tD^ke + e^tE^kAx_1^k)} && (6)\\
\Delta\tilde{x}^k &= \tilde{x}_0^k + \Delta\tilde{\beta}^k x_1^k && (7)\\
\Delta\tilde{u}^k &= e\Delta\tilde{\beta}^k + A\Delta\tilde{x}^k && (8)\\
\Delta\tilde{v}^k &= e\Delta\tilde{\beta}^k - A\Delta\tilde{x}^k && (9)\\
\Delta\tilde{y}^k &= -y^k - Y^k(U^k)^{-1}\Delta\tilde{u}^k && (10)\\
\Delta\tilde{z}^k &= -z^k - Z^k(V^k)^{-1}\Delta\tilde{v}^k && (11)\\
x_0^k &= (A^tD^kA)^{-1}A^t((\mu^k((U^k)^{-1} - (V^k)^{-1})e + z^k - y^k - \\
&\quad (U^k)^{-1}\Delta\tilde{u}^k\Delta\tilde{y}^k + (V^k)^{-1}\Delta\tilde{v}^k\Delta\tilde{z}^k) && (12)\\
\Delta\beta^k &= \frac{(e^t((\mu^k((U^k)^{-1} + (V^k)^{-1})e - z^k - y^k) - e^tE^kAx_0^k)}{(e^tD^ke + e^tE^kAx_1^k)} && (13)\\
\Delta x^k &= x_0^k + \Delta\beta^k x_1^k && (14)\\
\Delta u^k &= e\Delta\beta^k + A\Delta x^k && (15)\\
\Delta v^k &= e\Delta\beta^k - A\Delta x^k && (16)\\
\Delta y^k &= (\mu^kI + \Delta\tilde{Y}^k\Delta\tilde{U}^k)(U^k)^{-1}e - y^k + Y^k(U^k)^{-1}\Delta u^k && (17)\\
\Delta z^k &= (\mu^kI - \Delta\tilde{Z}^k\Delta\tilde{V}^k)(V^k)^{-1}e - z^k - Z^k(V^k)^{-1}\Delta v^k && (18)\\
\hat{\alpha}_p^k &= \min_i\left\{\min\left\{\frac{-u_i^k}{\Delta u_i^k}\,|\Delta u_i^k < 0\,;\frac{-v_i^k}{\Delta v_i^k}\,|\Delta v_i^k < 0\right\}\right\} && (19)\\
\hat{\alpha}_d^k &= \min_i\left\{\min\left\{\frac{-y_i^k}{\Delta y_i^k}\,|\Delta y_i^k < 0\,;\frac{-z_i^k}{\Delta z_i^k}\,|\Delta z_i^k < 0\right\}\right\} && (20)\\
(\alpha_p^k, \alpha_d^k) &= \begin{cases} (\hat{\alpha}_p^k, \hat{\alpha}_p^k), \text{ if } \hat{\alpha}_d^k > \hat{\alpha}_p^k \text{ and } y_k^t\Delta u^k + z_k^t\Delta v^k > 0,\\ (\hat{\alpha}_d^k, \hat{\alpha}_d^k), \text{ if } \hat{\alpha}_p^k > \hat{\alpha}_d^k \text{ and } u_k^t\Delta y^k + v_k^t\Delta z^k > 0,\\ (\hat{\alpha}_p^k, \hat{\alpha}_d^k), \text{ otherwise.}\end{cases} && (21)\\
w^{k+1} &= w^k + \tau^k\alpha_p^k\Delta w^k && (22)\\
y^{k+1} &= y^k + \tau^k\alpha_d^k\Delta y^k && (23)\\
z^{k+1} &= z^k + \tau^k\alpha_d^k\Delta z^k && (24)
\end{aligned}
$$

Until Convergence.

Figure 4: Predictor-Corrector Method (PC)

Given x^0, $r^0 = b - Ax^0$ and y^0 such that $A^t y^0 = 0$ and $e^t y^0 < 1$.
For $k = 0, 1, 2, \ldots$, do

$$
\begin{aligned}
|\rho_j^k| &= \|r^k\|_\infty && (1)\\
s^k &= sign(r^k) && (2)\\
g^k &= \sigma_j^k e_j && (3)\\
T^k &= [-\sigma_1^k e_1, \cdots, -\sigma_{j-1}^k e_{j-1}, s^k, -\sigma_{j+1}^k e_{j+1}, \cdots, -\sigma_n^k e_n] && (4)\\
f^k &= (T^k)^{-1} r^k && (5)\\
\beta^k &= \frac{\|F^k (T^k)^t (g^k - y^k)\|}{\|r^0\|_\infty} + \| \max_{i=1,\cdots,n} (-\sigma_i^k \lambda_i^k, 0)\|_\infty && (6)\\
\theta^k &= \frac{\beta^k}{(\nu + \beta^k)} && (7)\\
D^k &= F^k diag((1-\theta^k)|(T^k)^t (g^k - y^k)| + \theta e)^{-1} && (8)\\
\Delta x^k &= (A^t (T^k)^{-t} D^k (T^k)^{-1} A)^{-1} A^t g^k && (9)\\
\Delta r^k &= -A \Delta x^k && (10)\\
y^{k+1} &= g^k + (T^k)^{-t} D^k (T^k)^{-1} \Delta r^k && (11)\\
\alpha_*^k &= \max_i \left\{ \alpha_i^k \,\middle|\, \sigma_i^{k+1} \Delta \rho_i^k < 0 \right\} && (12)\\
\alpha_l^k &= \min_i \left\{ \alpha_i^k \,|\, \alpha_i^k > \alpha_*^k \right\} && (13)\\
\alpha^k &= \alpha_*^k + \tau(\alpha_l^k - \alpha_*^k) && (14)\\
x^{k+1} &= x^k + \tau \alpha^k \Delta x^k && (15)\\
r^{k+1} &= r^k + \tau \alpha^k \Delta r^k && (16)
\end{aligned}
$$

Until Convergence.

Figure 5: Coleman-Li Method (C&L)

Section 5. The convergence criteria and starting points are also presented later.

Because we noticed some convergence difficulties when $\mu^k = \left(\frac{\gamma^k}{2n}\right)^2$ in Step (1) of the methods shown in Figures 3 and 4, as proposed by Zhang, this perturbation parameter was changed to $\mu^k = \left(\frac{\gamma^k}{(2n)^{\frac{3}{2}}}\right)$.

The line search procedure to compute the step length in equation (12) of the C&L method (shown in Figure 5) is computed with the following parameters (α_i^k):

$$\left\{ \alpha_i^k > 0 \,\middle|\, \alpha_i^k = -\frac{(|\rho_j^k| - |\rho_i^k|)}{(\sigma_j^k \Delta\rho_j^k - \sigma_i^k \Delta\rho_i^k)} \text{ or } -\frac{(|\rho_j^k| + |\rho_i^k|)}{(\sigma_j^k \Delta\rho_j^k + \sigma_i^k \Delta\rho_i^k)} \right\}.$$

As opposed to the $O(n)$ line search usually used in interior point methods, this procedure is $O(n^2)$ in the worst case, since it is necessary to compute the new residuals for σ_i^{k+1}. However, when $\theta^k \leq 0.001$, Coleman and Li suggest to take the first positive step length anyway, thus avoiding the worst case scenario close to a solution. In this situation, $\alpha_* = 0$ and $\alpha_l = \min_i\{\alpha_i > 0\}$ in Figure 5, Step (14). Finally, the computation of D^{k+1} changes slightly whenever $\alpha_*^k = 0$ and

$$|\sigma_l^k \lambda_l^{k+1} - \sigma_h^k \lambda_h^{k+1}| < (1-\tau)\sigma_l^k \lambda_l^{k+1}$$

where $\alpha_h^k = \min_i \left\{\alpha_i^k \,|\alpha_i^k > \alpha_l^k\right\}$. If that happens, $\frac{1}{2}\theta^{k+1}$ is subtracted from δ_{ll}^{k+1}.

4 Enhancements to the Polynomial L_∞ Fitting Problem

4.1 Computational Requirements of Interior Point Methods

The most expensive step for all interior point methods is the solution of a linear system with the matrix $A^t D^* A$, where D^* is a diagonal matrix whose diagonal entries depend upon the method. For the C&L method, however, D^* is premultiplied by the elementary matrix T^{-t} and postmultiplied by its transpose. This feature brings extra overhead for the C&L method. Nonetheless, the linear system remains symmetric definite positive.

The prevalent path to solve the linear system is to compute the matrix A^tD^*A and its Cholesky factorization. As D^* is already known, computing the matrix involves $\frac{3nm(m+1)}{2}$ flops (flop is an acronym for float point operation). Cholesky factorization requires about $\frac{m^3}{3}$ flops and the solution of each triangular system is achieved with $2m^2 - m$ flops. Multiplications for computing the matrix can be halved by storing the entries of the dot products for each pair of columns of A; however, this approach requires $\frac{nm(m+1)}{2}$ additional positions of storage. All methods require at least one matrix-vector product, involving A and A^t; each matrix-vector product requires $2mn$ flops.

Taking all operations into account, the time computational complexity of one iteration is $O(nm^2)$ for all L_∞ interior point methods and $O(nm^2 + n^2)$ for C&L (recall that $m \leq n$). The number of linear systems to complete the iteration varies with the method. The primal method solves only one linear system. The predictor-corrector variant requires the solution of four linear systems. Each of the other methods requires two linear system solutions. The Cholesky factorization from the first system can be used for solving the remaining ones in the iteration, since all share the same matrix. Thus, solving each extra triangular system requires only $2m^2 - m$ additional flops.

Note that we did not have any concern about sparse structure in this discussion. They are indeed unnecessary, since the constraint matrix for the L_∞ fitting problem is usually full.

4.2 Specialization to Polynomial L_∞ Fitting Problems

In Oliveira et al. (2000) the special case where the data is approximated in the L_1 sense by a polynomial of degree $d = m - 1$ was considered. In this situation, the constraint matrix has a very structured form,

$$V^t = \begin{pmatrix} 1 & 1 & \cdots & 1 \\ \alpha_1 & \alpha_2 & \cdots & \alpha_n \\ \vdots & \vdots & & \vdots \\ \alpha_1^{m-1} & \alpha_2^{m-1} & \cdots & \alpha_n^{m-1} \end{pmatrix}$$

which is called a rectangular *Vandermonde matrix*. Moreover, the product $H = VD^*V^t$ has the special structure of a *Hankel matrix*; its entries are constant along its antidiagonals. The same happens for the interior point methods applied to the L_∞ polynomial fitting problem.

With a careful use of this structure, the computational complexity of an iteration can be reduced. The reason is that there exist $O(m^2)$ algorithms for solving Hankel systems (e.g., Trench, 1965) and H can be obtained in $O(nm)$ flops, since only $2m-1$ entries are computed.

Also, there is no need to store any matrix in all the interior point methods studied, due to the special structure of V. The matrix-vector products Vx or $V^t y$ can be computed with the usual amount of flops without storing V. The same applies for computing H from V. Also, since H has only $2m-1$ different entry values, it can be stored as a vector. Furthermore, it is possible to compute its inverse in $O(m^2)$ flops (Trench, 1965) and simultaneously solve the linear system. Thus, no matrix needs to be stored for any procedure. Notice that this discussion does not completely apply for the C&L method since, in this case, $H = VT^{-t}DT^{-1}V^t$ is not a Hankel matrix.

Although there is no need to store matrices, it is convenient to do so in all methods, except for the primal. When the solution of extra linear systems are necessary to accomplish the iteration, having H^{-1} at hand will speed up the method.

There is yet another improvement for the polynomial fitting problem. It is easy to find a basis from an interior point without storing any matrix, as well. In order to compute it, it is enough to select the m smallest residuals $|r_i|$ and interpolate the corresponding points. As the values of α_i are all distinct, any set of columns of V will lead to a nonsingular matrix for the interpolation. After the interpolation, a simplex based code can be used to refine the solution, if desirable.

5 Numerical Experiments

Numerical experiments were carried out to compare the standard and improved versions of interior point approaches to L_∞ fitting problems against the time-proved specialized simplex approach proposed by Barrodale and Phillips (1975) (B&P).

The simplex method implementation is the same that appears in the original paper from Barrodale and Phillips coded in FORTRAN, with few increases in dimensions of arrays to enable solution of larger problems. Interior point methods were coded by the authors in C. All case studies were run on a Sun Ultra1 workstation with double precision floating point arithmetic.

5.1 Parameters

The main parameters were defined as follows. Stopping criteria were based on the optimality conditions. It measures the relative dual infeasibility and the relative gap:

$$\begin{pmatrix} \frac{\|A^t(y^k-z^k)\|}{1+\|y^k\|+\|z\|+\|a\|} \\ \frac{|e^t(y^k+z^k)-1|}{1+\|y^k\|+\|z\|+\|a\|} \end{pmatrix} \leq \epsilon, \quad \frac{\max_i\{\max(u_i^k y_i^k; v_i^k z_i^k)\}}{1+\gamma^k} \leq \epsilon,$$

where $a = (\alpha_1, \ldots, \alpha_n)$. Thus, $\|a\|$ is an estimate of $\|A\|$. The stopping tolerance ϵ is the square root of machine precision (Dennis and Schnabel, 1996).

Starting points were based on Zhang (1993): $x^0 = (A^tA)^{-1}A^tb$, $r^0 = b - Ax^0$, $\beta^0 = \lambda\|r^0\|_\infty$, $u^0 = e\beta^0 - r^0$, $v^0 = e\beta^0 + r^0$, $y^0 = z^0 = \frac{e}{2n}$.

In all interior point methods, $\lambda = 1.1$ and τ was set to 0.95. In C&L the dual variable y starts with $y^0 = \kappa\frac{r^0}{\|r^0\|_\infty}$ where $\kappa = 0.975$, as used by Coleman and Li (1992), and the parameter ν in Step (7) of Figure 5 was set to 0.99.

5.2 Case Studies

Numerical experiments were designed to compare CPU time required by each of the methods as a function of the number of observations (n) and degree of the fitting polynomial ($d = m - 1$). We borrowed examples from a previous paper on the L_1 fitting problem (Oliveira et al., 2000), where several data sets were randomly generated with the following features:

- n was fixed at 10000 and d set at values 1, 2, 3 and 5 (results from these data sets are shown in Figure 6);

- d was fixed at the value 3 and n set at 1000, 5000, 10000 and 20000 (results with experiments from these data sets are shown in Figure 7).

The values adopted for the degree of the fitting polynomials (d) were small because we are not aware of real-life applications which use curves of high degrees. Values for the independent variables are n equally spaced points in the range [0,1]. Dependent variables were randomly generated with uniform U(0,1), normal N(0,1) and exponential Exp(1) distributions. Since the results obtained with different distributions had the same qualitative behavior, only results from the U(0,1) distribution are shown. Results presented in Figures

6 and 7 are the average obtained from several runs of the codes—input and output times were not accounted. In order to make comparisons easier, all curves have the same y-axis range, except for C&L because it was slower than the other methods. The performance of the simplex method was used as a yardstick.

From now on the extensions "s" which goes for the standard (slow) version and "f" for the specialized (fast) version are added to the acronym of interior point methods proposed in this paper. For instance, the Primal method versions are represented as P-s and P-f.

Observation of Figures 6 and 7 reveals the simplex method as a clear winner in solving L_∞ problems. This somewhat unexpected result stems from the small number of iterations required by the simplex approach. Iterations of interior point methods are more elaborated and require more computation than iterations of simplex. However, interior point approaches typically beat simplex in the solution of large linear programming problems, because they usually require much fewer iterations. When the number of simplex iterations to solve a problem is small, it is hard to find interior point directions that will be good enough to compensate for it.

From Figure 6, we observe that the specialized versions of interior point methods outperform the standard versions only when the degree of the approximate polynomial increases—in the primal the specialized version shows its qualities at a lower degree, but the general trend is the same. The standard C&L version, however, behaves better than the specialized one for all the fitting polynomial problems studied because of the structure of the problem cannot be completely exploited in the specialized version of C&L approach.

Figure 7 shows that the specialized versions beat the standard only for the primal. The reason seems to be that, as opposed to all other methods, the primal solves only one linear system for computing the search directions allowing its solution without storing any matrix.

The number of iterations required by each method is summarized on Table 1 (averages were obtained for all numerical experiments with randomly generated data). In most cases, standard and specialized implementations yielded the same number of iterations. Once again C&L was the exception; large differences in the number of iterations between standard and specialized versions were observed. Whenever there was such a difference, we used the highest value between the standard and specialized versions to compute averages registered in Table 1. For two instances primal-dual (PD) presented numerical problems and failed to achieve convergence.

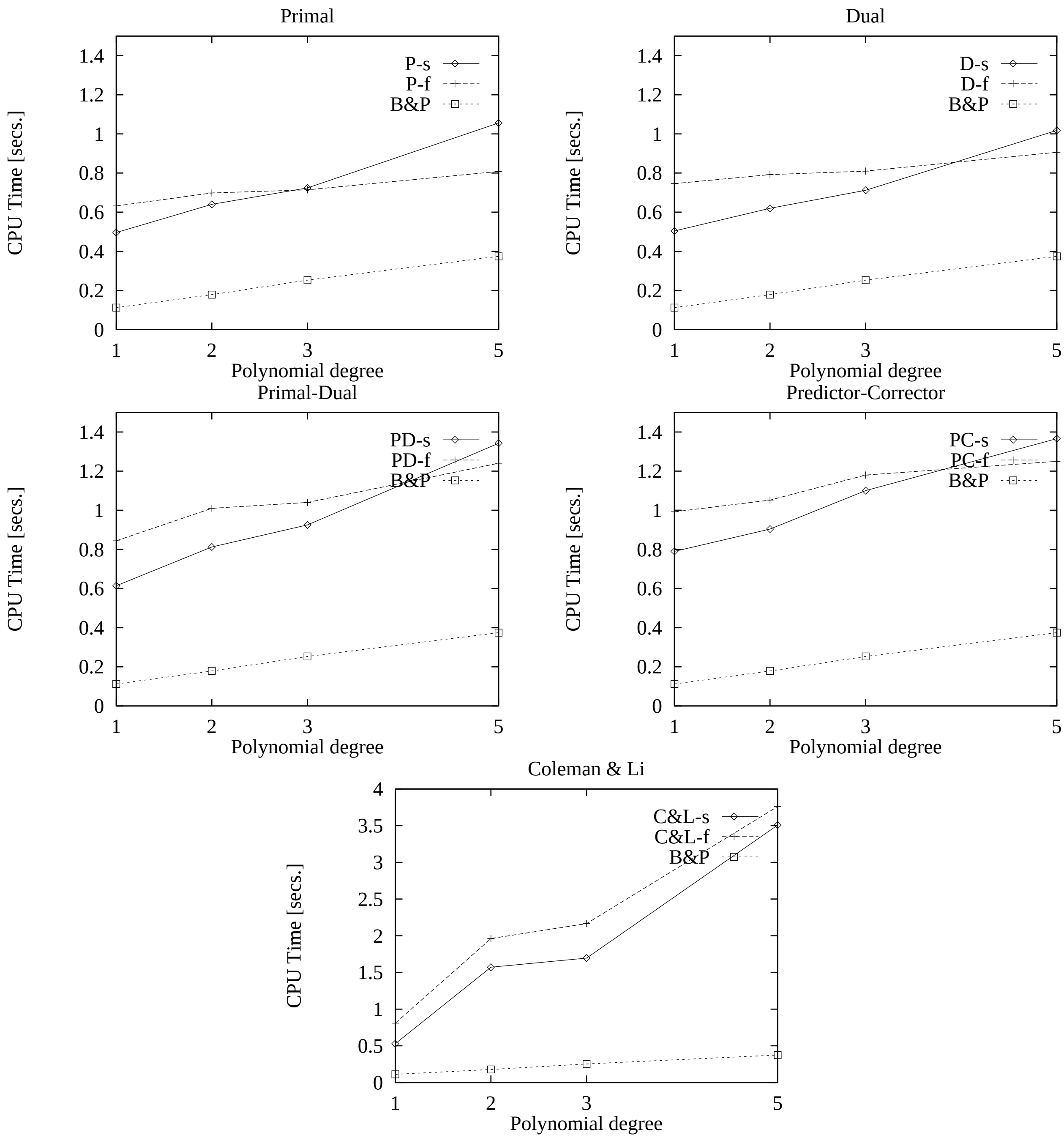

Figure 6: CPU Time as a Function of the Degree

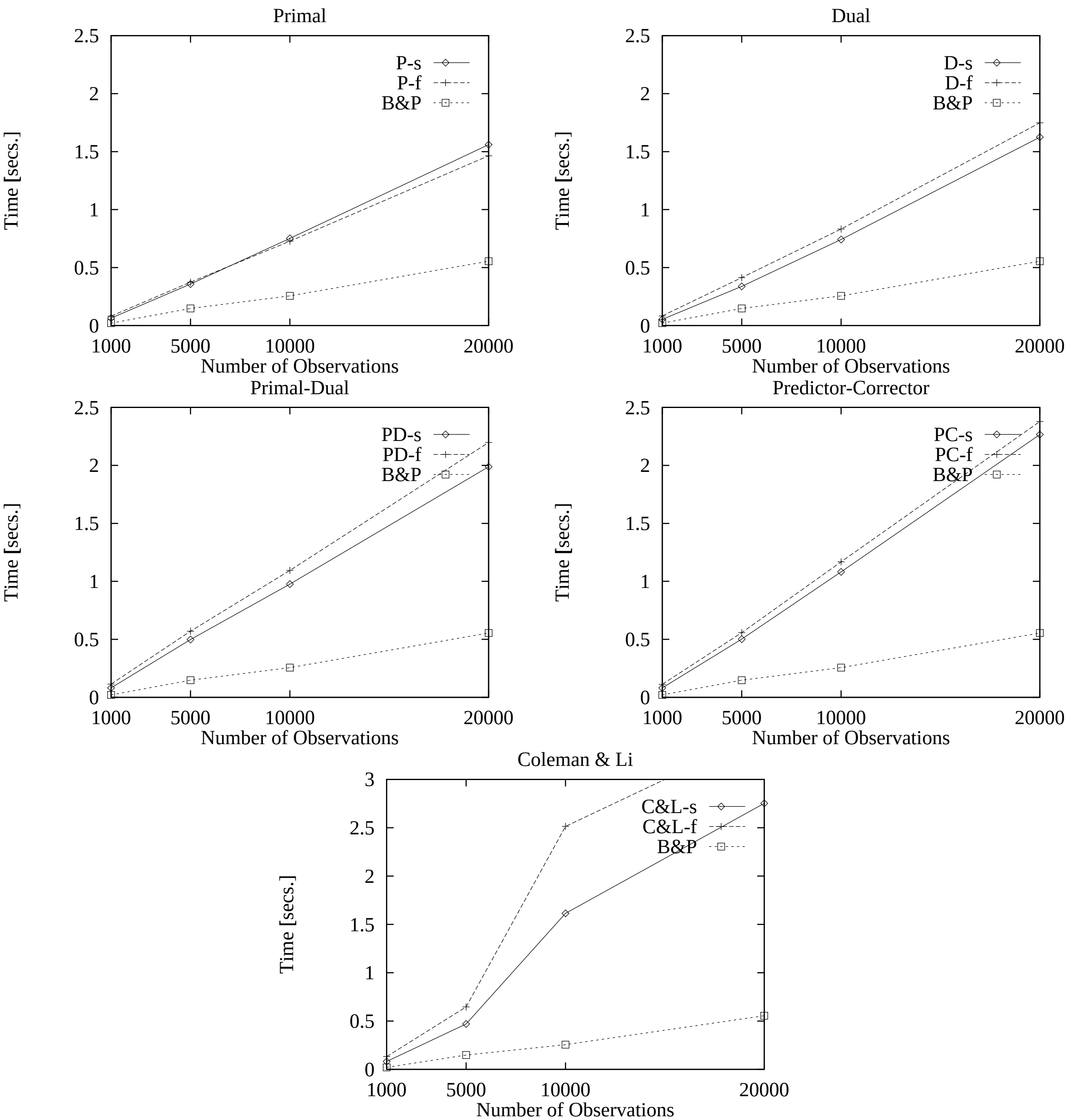

Figure 7: CPU Time as a Function of the Number of Observations

Note that the simplex method presents the smallest average in the number of iterations, leaving no room for better performance of interior point methods. Table 1 also shows that the predictor-corrector (PC) interior point method performs very well with regard to the number of iterations; in other words, interior point directions for PC are better. However, it is not as fast as other interior point methods, since it requires the solution of four linear systems to compute a direction. Finally, examination of Table 1 shows that the average number of iterations for C&L is not very informative, since the maximum and minimum number of iterations are far apart.

Table 1: Number of Iterations Required by the IPMs

	Average	Minimum	Maximum
P	15.8	14	18
D	16.8	14	20
PD	16.7	13	23
PC	12.6	10	16
C&L	23.4	7	88
B&P	11.5	4	21

We also run two experiments using large sets of real financial data. The first set contains the "prime rate" (PR) data. The second set contains the "federal fund rate" (FFR) data. Both data sets comprise daily values for over 40 years, weekends and holidays not considered, totaling 10958 observed values. To improve numerical stability of the methods, data (observations) were normalized in the range [0,1]. Nevertheless, the primal-dual (PD) and predictor-corrector (PC) did not converge for cases with fitting polynomials of degree 5.

Results obtained are shown in Table 2 (a and b), where computation times of interior point methods are for the fastest version (either standard or specialized) – once again, the specialized primal version was often faster than the standard one, whereas the opposite occurred for the other methods. These experiments confirm that B&P simplex based algorithm presents the best overall time performance. Among the interior point methods, the primal affine-scaling proposed in this paper shows more stable results; also, it was the only one to beat B&P (for data set PR with fitting polynomial of degree 4). For most problems, interior point methods and simplex provided the

Table 2: CPU Time [secs.] as a Function of the Degree d

(a) Data set PR (b) Data set FFR

d =	1	2	3	4	5	1	2	3	4	5
P	0.64	0.69	1.32	0.84	1.33	0.62	1.10	1.37	1.11	1.15
D	0.47	0.86	0.91	1.00	1.29	0.56	1.37	1.87	2.39	1.67
PD	0.64	1.57	2.11	1.90	-	0.68	1.82	2.57	1.33	1.51
PC	0.94	1.35	1.34	1.63	1.58	1.00	1.27	1.61	2.18	-
C&L	0.80	1.71	2.34	2.19	2.35	0.80	0.99	2.56	1.50	0.85
B&P	0.12	0.31	0.46	0.93	0.87	0.15	0.25	0.36	0.41	0.42

same values for the objective function, within eight digits.

6 Conclusions

This paper compared interior point methods to L_∞ fitting problem with the simplex based approach developed by Barrodale and Phillips. All main interior point approaches were reviewed: dual affine-scaling, primal-dual and its predictor-corrector variant. The method proposed by Coleman and Li closely related to the interior point family, was also reviewed. A primal affine-scaling method is proposed, completing the affine-scaling interior point alternatives for L_∞ fitting problems.

All interior point methods were specialized to take advantage of the structure present in polynomial fitting problems. For this class of problems it was shown that there is no need to store matrices and computational complexity was reduced.

Extensive numerical experiments pointed to the Barrodale and Phillips simplex based method as a clear winner in the match of approaches to L_∞ fitting problems. It is interesting to notice that the interior point approaches to L_1 problems leaded to completely different results (Oliveira et al., 2000). This class of problems shares many properties with the L_∞ class. Yet, our implementation was able to beat a classical specialized simplex implementation in similar conditions.

The specialized primal affine-scaling method performed better among the interior-point approaches; it is more stable and was the only one to beat simplex in an instance of polynomial fitting. This achievement of the primal affine-scaling interior point method may seem surprising, as the predictor-corrector is considered the best alternative for general large linear programming problems. However, the primal affine-scaling takes better benefits from the special structure of the L_∞ fitting problems—similar features are shared by the dual affine-scaling approach for the L_1 fitting problem (Oliveira et al., 2000). The primal affine-scaling is the interior point method with smallest overall effort to accomplish an iteration; it requires the solution of only one linear system at each iteration and imposes less additional overhead. Also, it allows easy choices of good starting points.

The specialized versions of the interior point methods did not present steady better time performance as it happen with their counterparts for L_1 fittings. However, they became more competitive as fitting degrees of polynomials increase. The primal affine-scaling method, however, was a exception. It exhibited better results for the specialized version, probably because it solves only one linear system per iteration and avoids the storage of any matrix .

To conclude, we need to address the key point of which approach to recommend for dealing with L_∞ fitting problems. As a clear winner in our numerical experiments, the simplex approach should be the first choice if time performance is the major concern.

On the other hand, note that in our numerical experiments all methods considered achieved L_∞ fitting within a few seconds. Since interior point methods for the L_∞ problem are easy to code and building the adapted simplex method requires much more work, this can be another facet to take into account. The primal affine-scaling method, in addition to present the best performance among interior point approaches, is the easiest one to implement.

Acknowledgments

This research was partially sponsored by Foundation for the Support of Research of the State of São Paulo, (FAPESP) and the Brazilian Council for the Development of Science and Technology (CNPq). Thanks also go to Mário Nascimento for cooperation in early stages of this research.

References

Arthanari, T. S. and Dodge, Y. (1981). *Mathematical Programming in Statistics*, John Wiley, New York.

Barrodale, I. and Phillips, C. (1975). Algorithm 495: Solution of an overdetermined system of linear equations in the Chebychev norm, *ACM Trans. on Mathematical Software* **1**: 264–270.

Coleman, T. F. and Li, Y. (1992). A global and quadratically convergent method for linear l_∞ pr oblems, *SIAM J. Numer. Anal.* **29**(4): 1166–1186.

Dennis, J. E. and Schnabel, R. B. (1996). *Numerical Methods for Unconstrained Optimization and Nonlinear Equations*, SIAM, Philadelphia, PA.

Karmarkar, N. (1984). A new polynomial-time algorithm for linear programming, *Combinatorica* **4**(4): 373–395.

Narula, S. C. and Wellington, J. F. (1988). An efficient algorithm for the MSAE and MMAE regression problems, *SIAM J. Sci. Stat. Comput.* **9**(4): 245–253.

Oliveira, A. R. L., Nascimento, M. A. and Lyra, C. (2000). Efficient implementation and benchmark of interior point methods for the polynomial L_1 fitting problem, *Statistics & Data Analysis* **35**(2): 119–135.

Ruzinsky, S. A. and Olsen, E. T. (1989). l_1 and l_∞ minimization via a variant of Karmarkar's algorithm, *IEEE Transactions on Acoustics, Speech, and Signal Processing* **37**(2): 245–253.

Trench, W. F. (1965). An algorithm for the inversion of finite Hankel matrices, *J. Soc. Indust. Appl. Math.* **13**(4): 1102–1107.

Wagner, W. H. (1959). Linear programming techniques for regression analysis, *J. Amer. Statist. Assoc.* **54**: 206–212.

Zhang, Y. (1993). A primal-dual interior point approach for computing l_1 and l_∞ solutions of overdetermined linear systems, *Journal of Optimization Theory and Applications* **77**(2): 323–341.