THE ASTROPHYSICAL JOURNAL, 398:476–490, 1992 October 20

AN OBJECTIVE CLASSIFICATION SCHEME FOR QSO SPECTRA[1]

PAUL J. FRANCIS[2] AND PAUL C. HEWETT
Institute of Astronomy, Madingley Road, Cambridge CB3 0HA, England

AND

CRAIG B. FOLTZ AND FREDERIC H. CHAFFEE[3]
Multiple Mirror Telescope Observatory, University of Arizona, Tucson, AZ 85721
Received 1992 February 18; accepted 1992 April 24

ABSTRACT

A new approach to the classification of QSO rest-frame optical and ultraviolet spectra is presented. The statistical technique of principal component analysis (PCA) is applied directly to a sample of QSO ultraviolet spectra, rest-frame wavelengths $\lambda\lambda$1150–2000, taken from the Large, Bright QSO Survey. A detailed discussion of the application of the PCA technique to QSO spectra is given. The PCA approach provides new insights into the relation between the continuum, emission-line, and broad-absorption-line properties of QSOs. The QSO spectra can be described using a small number of principal components, and a quantitative classification scheme for the rest-frame ultraviolet region has been devised. The major results of the analysis can be summarized as follows:

1. The first three principal components obtained from the PCA account for ~75% of the QSO sample intrinsic variance, and a quantitative classification scheme based on the contribution of the components to each QSO spectrum can be developed.
2. One of the components describes the emission-line core strength, the second the continuum slope, and the third the broad absorption features present in a subset of the spectra.
3. Broad-absorption-line QSOs are naturally identified as a subclass of the QSO population. They show excess N v and Al III + C III] emission relative to the QSO population as a whole.
4. The emission lines of the QSO sample show a spread of properties which can be arranged in a sequence ranging from broad, low-equivalent-width lines to narrow, high-equivalent-width lines. Other properties of the QSO spectra also vary along this sequence. For example, QSOs in the former category have higher N v/Lyman α, N v/C IV, and C IV/Lyman-α line ratios than objects with lines characteristic of the latter category, and the wavelength separation between the Lyman-α and C IV emission lines is larger in spectra in the former category than in the latter.
5. The continuum hardness correlates with the equivalent widths of C IV and Lyman-α emission lines but with a large scatter, and the equivalent width of the Al III/C III] blend correlate with a smaller scatter.
6. A Baldwin Effect consistent with the results of other studies is present in the LBQS sample.

Subject headings: quasars: general — ultraviolet: galaxies

1. INTRODUCTION

The classification of data involves the attempt to distill some subset of observed phenomena to a small number of common characteristics. In so doing, some hitherto unrecognized pattern may emerge that will provide a clue to the underlying nature of a particular object or population of objects. In astronomy, perhaps the best known example is the MK system of stellar spectral classification which has provided a framework for insight into the physical nature of stars. Nearly 8000 QSOs have now been identified since the discovery of QSOs nearly 30 years ago, and they display a wide range of spectral characteristics in both their continuum and emission-line properties. Despite this wealth of information, no simple classification scheme has emerged to describe the rest-frame optical and ultraviolet spectra of these important objects. Progress toward a classification scheme has been confined to phenomenological description of certain characteristics of certain QSO subclasses (e.g., Sulentic 1989).

The failure to identify unifying characteristics among QSO spectra has meant the analyses of samples of QSOs have provided few quantitative constraints on theoretical models of the QSO phenomenon. The Baldwin effect (Baldwin 1977), a weak anticorrelation of the integrated flux of the C IV emission line with continuum luminosity, remains the principle statistical relation to be established between spectroscopic properties which may reflect a connection between aspects of theoretical models, in this case the continuum radiation source and the broad-line region.

The slow progress toward identifying unifying characteristics can be attributed to two factors: (1) the lack of a large sample of QSOs containing the full range of spectral properties exhibited by the population for which homogeneous spectroscopic observations to an adequate signal-to-noise ratio (S/N) are available, and (2) the adoption of an analysis methodology whereby many discrete features are identified in a sample of QSO spectra prior to the examination of possible correlations

[1] Observations reported here were obtained with the Multiple Mirror Telescope, a facility operated jointly by the Smithsonian Institution and the University of Arizona, and with the Las Campanas Observatory, a facility of the Carnegie Institution of Washington.

[2] Present address: Steward Observatory, University of Arizona, Tucson, AZ 85721.

[3] Senior Visiting Fellow, Institute of Astronomy, 1990 July 1–1991 August 1.

and relationships between these features. The decomposition of emission lines into a series of symmetric profiles with different velocity widths, or the description of the continuum shape in the rest-frame ultraviolet as some combination of power-law and thermal components are common examples of such an approach.

This paper explores a new application of principal component analysis (PCA) in which the algorithm is applied directly to QSO spectra rather than to quantities measured from the spectra. The goal is to identify underlying spectral characteristics responsible for the extensive variations in properties observed in QSO spectra. The criterion by which the technique will be judged is whether the rest-frame ultraviolet spectra of a large sample of QSOs can be described by a relatively small number of "components."

The investigation employs spectra from the Large, Bright QSO Survey (LBQS), whose main properties are described in § 2. Section 3 contains a discussion of the PCA technique and details of its application to the sample. The results follow in § 4 where the significant components resulting from the PCA analysis are described and a simple classification scheme based upon the strongest components is presented. Finally, § 5 summarizes the main conclusions.

2. THE LARGE BRIGHT QSO SURVEY

The LBQS consists of data on 1018 QSOs, with redshifts $0.2 \leq z < 3.4$, and apparent magnitudes $16.0 \leq m_{B_J} \leq 18.85$. A description of the observing procedures and plots of all the spectra can be found in Foltz et al. (1987, 1989), Hewett et al. (1991), Chaffee et al. (1991), and Morris et al. (1991). The LBQS candidate selection criteria were defined to be sensitive to all objects that might reasonably be described as QSOs (detected by any means), provided only that the observed flux in the B_J passband places the objects within the survey apparent magnitude limits. The sole exception to this QSOs whose spectra are devoid of emission features, absorption features, and continuum breaks, and have spectral energy distributions redder than that of a power law of slope $\alpha \sim -2.0$ in the observed-frame wavelength range $\lambda\lambda$3500–5200. Therefore, with the exception of red, featureless BL Lac objects the LBQS is sensitive to all types of confirmed "QSOs" of which we are aware.

The calculation of the LBQS selection function—the probability of identifying a QSO of specific spectral properties—is not yet complete, and the relative numbers of QSOs with a specific type of spectrum making up the population as a whole has not been determined. Nonetheless, the critical factor for the development of a classification scheme is that the LBQS sample contains QSOs representing the full range of spectroscopic properties present in the population.

The LBQS spectra were obtained using the Multiple Mirror Telescope (MMT), whose data provide a spectral coverage of $\lambda\lambda$3200–7400 at 6 Å resolution, or with the Du Pont Telescope of the Las Campanas Observatory (LCO) ($\lambda\lambda$3200–6800 at 10 Å resolution). Each QSO was observed until a S/N of $\sim$10 was obtained in the continuum region nearest to 4000 Å in the observed frame. This procedure ensured that the data are sufficiently homogeneous that differences in S/N or resolution do not significantly affect the analysis.

3. PRINCIPAL COMPONENT ANALYSIS AND QSO SPECTRA

3.1. *Principal Component Analysis*

PCA is a mathematical tool with the potential to simplify multidimensional data sets. Its application to data described by n variables results in a new description of the data consisting of n new variables[4] each consisting of a linear combination of the original variables. In cases where a smaller number of the new variables can account for the bulk of the total variance present in the data, the effective "dimensionality" has been reduced, and the first goal of any classification scheme achieved. PCA is widely used in the social sciences (see Kendall 1975 for an overview and examples from many fields) and has been applied several times in astronomy (see Murtagh & Heck 1987 and the many references cited therein, and Whitney 1983b). Recently Boroson & Green (1992) have applied PCA to spectra of low-redshift QSOs from the Palomar Green survey. Their approach was to use the technique to look for correlations between measured properties of the spectra (e.g., line strengths and line profiles). In the way it is employed here, PCA has the advantage of elucidating potential correlations *without* parameterizing the spectra by, for example, establishing a continuum level.

A description of the mathematical formalism together with a number of technical aspects of PCA can also be found in Whitney (1983a) and Mittaz, Penston, & Snijders (1990). In this section a qualitative description of the procedure is presented, and only aspects directly related to the analysis of QSO spectra are considered in detail.

Consider each QSO rest-frame spectrum as a point in a multidimensional vector space, the axes of which represent the fluxes in the individual wavelength bins. (All of the analysis is carried out with spectra normalized to have a unit mean flux over the wavelength range of interest. This ensures that the analysis is sensitive only to variations in the *shape* of the spectra.) The number of dimensions is determined by the number of resolution elements within the wavelength range under investigation. For the rest-frame wavelength region and instrumental resolution employed here, 284 resolution elements, each 3 Å wide, are used. The data form a cloud of points within this multidimensional space. The greater the range of fluxes from QSO to QSO in a given rest-wavelength bin, the more extended the points will be along the corresponding axis. Thus, a bin situated at the peak of an emission line is likely to show a greater range of values than one situated in a continuum region. If the fluxes in two or more bins are correlated, the cloud points will be elongated along some direction in the plane defined by their two axes, and the cloud extension will be roughly proportional to the square root of the number of correlating bins. The directions along which the cloud of points are most elongated are of interest as they represent the dominant sources of variation among QSO spectra. Large extensions can arise when one, or a few correlated, wavelength bins exhibit a large range of values, or when smaller correlated variations occur over a substantial number of wavelength bins.

PCA identifies these directions and uses them as a set of axes for a new parameterization of the multidimensional space. Following the analysis each QSO spectrum can be represented by its coordinates in this new parameterization. The new axes are identified sequentially; PCA first finds the most extended direction in the original n-dimensional space by choosing that from which the sum of the squared deviations of the points is a minimum. This direction forms the "first principal component" and accounts for the largest single linear variation from one QSO spectrum to another. All the points are then collapsed down to the $(n-1)$–dimensional space perpen-

[4] If the number of objects m for which data are available is less than n, only m new variables will be found.

dicular to this axis, thus removing this largest source of variation. The process is repeated, identifying higher-order principal components at each step, until the number of points (QSO spectra) or dimensions (wavelength bins) is exhausted.

The mean QSO spectrum is subtracted from each individual spectrum prior to performing the PCA. This is equivalent to moving the origin of the multidimensional space to the center of the cloud of points.

While the application of PCA to n-dimensional data produces a description of the data in terms of n new dimensions, it does not itself necessarily lead to a simplified description of the data or offer any physical insight. If each QSO spectrum exhibited no correlations as a function of wavelength, and QSO spectra showed no correlations among themselves, then the spectra would be scattered along every direction in the multidimensional space, and PCA would require all n components to describe all the spectra, so little or nothing would be gained. If, however, there is some correlation within individual QSO spectra or between QSO spectra, large parts of this space may be empty, and fewer principal components might adequately describe any spectrum. Finally, where linear correlations are present, a relatively small number of new variables may account for a large fraction of the total sample variance, and a much more compact characterization of the data set has been achieved.

It is conventional to order the principal components according to the fraction of the total sample variance for which they account—with the first principal component accounting for the largest fraction, the second principal component the second largest, and so on. The fraction of the total sample variance accounted for by the first m components is a good measure of the utility of PCA. For example, a successful application of PCA to a sample of spectra could show that the majority of the sample variance is accounted for by the first few principal components—a dramatic reduction in the dimensionality of the data. Using PCA to classify QSO spectra thus offers the advantage of providing a description of the QSO population that employs only those principal components that contribute significantly to the variation in spectral properties.

Ideally the number of important principal components should be small, allowing QSOs to be parameterized (or classified) using a small number of variables. In the optimum case it should further be possible to interpret the form of each principal component and its varying strength from QSO to QSO in terms of characteristics of the QSOs themselves. In the remainder of this section several technical aspects of the PCA analysis are considered.

3.2. *Scaling the Spectra*

One of the problems in the application of PCA concerns the scaling of the input variables. A standard technique is to scale each variable to unit variance prior to the analysis. Such scaling ensures the analysis is not dominated by variables that exhibit the greatest absolute ranges. The concomitant disadvantage is that variables containing the majority of information are given the same weight as variables dominated by noise. The problem does not have an obvious solution because it is not known a priori which are the important variables.

The scaling of the input variables has no effect on the fundamentals of the PCA; the original n variables are always transformed into n principal components. Changing the scaling of the input variables simply produces principal components that are different linear combinations of the input variables—any correlations between variables are preserved independent of scaling, but the distribution of information among the principal components can change. In QSO spectra the range of fluxes at wavelengths corresponding to strong emission lines exceeds the range typical of the continuum regions, and scaling the fluxes to unit variance reduces the importance of the variation in the emission lines relative to variations in the continuum. As the continuum extends over many wavelength bins the strongest components resulting from PCA applied to unit-variance-scaled data are dominated by changes in the shape of the continuum among QSOs.

In order to investigate the effects of scaling, PCA was performed using three different weighting schemes: (1) no scaling, (2) flux values in each wavelength bin were scaled to unit variance, and (3) flux values for the sample in each wavelength bin were sorted by amplitude and the flux value in each spectrum replaced by its rank, eliminating the effect of any non-Gaussian distribution of values.

These experiments showed that, irrespective of the method of scaling, the bulk of the sample variance can be accounted for by the first 10 principal components, thus achieving the hoped-for reduction in the dimensionality of the data. Furthermore, the same 10 components are present in each case although the fraction of the total sample variance accounted for by each component differs depending on the weighting scheme. As the results of the PCA are insensitive to scaling, scheme (1) was adopted on grounds of simplicity and ease of interpretation, that is, fluxes are *unscaled.*

3.3. *Application of PCA to the QSO Spectra*

PCA was applied to a sample of 232 QSO spectra, selected such that their redshifts ($1.8 < z < 2.2$ for LCO spectra, $1.8 < z < 2.7$ for MMT spectra) place the rest-frame wavelength region $\lambda\lambda 1150$–2000, which contains the high-ionization lines of Lyman-α, N V, Si IV + O IV], C IV, and Al III + C III], in the observed optical window. The application is carried out in six steps:

1. The redshift of each QSO is determined by cross-correlation with an high S/N composite spectrum, produced as described by Francis et al. (1991). The spectra are then rebinned to a linear wavelength scale in the QSO rest frame. The wavelength increment, 3 Å, is chosen to match the instrumental resolution, but the results of the PCA are not sensitive to the exact choice of wavelength increment.
2. Each spectrum is normalized to unit mean flux over the rest-frame wavelength range to be analyzed. The normalized spectra are averaged to produce the mean QSO spectrum.
3. The QSO spectra are reduced to zero mean by subtracting the mean QSO spectrum.
4. PCA is applied using the singular-valued decomposition technique described by Mittaz et al. (1990). An analysis of 300 spectra each consisting of 200 wavelength bins takes approximately 30 CPU seconds on a VAX 3800.
5. PCA produces a series of principal components ordered by the fraction of the sample variance for which each accounts. The principal components are simple linear combinations of the input fluxes, and they may be illustrated graphically as "spectra."
6. Each observed QSO spectrum can now be reconstructed as a linear combination of a small number of principal components, the amplitude of each forming the basis for the classification scheme described in § 4.4.

3.4. *The Effects of Noise*

Two aspects of noise associated with the survey spectra can alter the principal components and their weights: shot noise due to the finite number of spectra, and the finite S/N of each spectrum.

The shot noise associated with the finite number of spectra is important because one is trying to define a multidimensional volume using a finite number of points (spectra), and structure smaller than the typical differences between neighboring survey spectra in the spectrum space cannot be measured. In addition, rare or extended subpopulations of QSOs may be poorly represented or be absent altogether. Two-hundred-and-thirty-two spectra are used to define the principal components presented in § 3.5, so QSO subpopulations which comprise $\lesssim 1\%$ of the parent population of LBQS QSO spectra may be poorly represented.

The finite S/N of the spectra can affect both the weights of the components in individual spectra and the components themselves. Since the S/N is not a very strong function of wavelength, each spectrum has a roughly spherical error array in the spectrum space, that is, its errors are of comparable sizes in all directions (wavelength bins). The weights of the components describing a spectrum (the spectrum's position in a classification scheme) define the nearest point on the classification hypersurface to the spectrum, that is, the projection of the spectrum onto the plane defined by the major principal components. Thus the error in the weight of each principal component is roughly equal to the typical error in the flux of each wavelength bin. For the LBQS spectra, the S/N per resolution element is ~ 10, the mean flux is normalized to unity, and hence the typical error in the weight of each principal component is ~ 0.1. Alternatively, note that the principal components are simply orthonormal linear combinations of the fluxes and hence should have similar errors. The first several principal components are, however, chosen for their large range of weights among the spectra, and hence the noise will be a smaller fraction of the total variance than it is for a particular wavelength bin.

The finite S/N also affects the form of the principal components, possibly introducing spurious features. The presence of noise can also alter the relation between two or more principal components of comparable strength, their relative importance being sensitive to the random fluctuations caused by the noise. In this case, information is being redistributed between components, but spurious features are not introduced. The size of artifacts introduced into the components by the finite S/N may be estimated by considering a principal component with rms weight l calculated from n spectra with a noise amplitude σ. The principal component lies along the most elongated direction of the cloud of points and can be approximated by a line joining the mean of the two half-samples with the lowest and highest component weights. Each of these means has a position uncertain by $\sigma/\sqrt{2}$, and their separation is $\sim 2l$. Therefore, since the principal components, by definition, have unit lengths, each bin in the principal component will have an associated uncertainty of order $\sigma/(l\sqrt{n})$.

A related consideration is the importance of the information contained in the higher-order principal components. It is demonstrated in § 3.6 that the first few principal components account for a large fraction of the sample variance. However, while individual higher-order components are dominated by noise, the possibility remains that the cumulative effect of principal components that are not used in any analysis may "undo" the results obtained using the first few principal components. This problem is discussed in § 3.6.2.

3.5. *Principal Components Derived for LBQS Spectra*

Figure 1 shows the mean QSO spectrum and the first six principal components derived as described in § 3.3.

The first component appears flat on the largest scales, but is strongly modulated at the positions of the familiar QSO emission lines, and thus can be considered as primarily a measure of emission-line strength. The second component reflects the variation in the slope of the continuum, but there are also modulations at the positions of the Lyman-α, N v, and C IV emission lines. The third component shows troughs to the blue of C IV, Si IV + O IV], and N v, with emission at N v, Si IV + O IV], C IV, and Al III + C III]. This component is a measure of the strength of the broad-absorption-line (BAL) phenomenon. The components are discussed in more detail in § 4. Higher-order components become increasingly difficult to characterize. They are nonetheless necessary to provide adequate reconstructions of some QSO spectra as described in the next section.

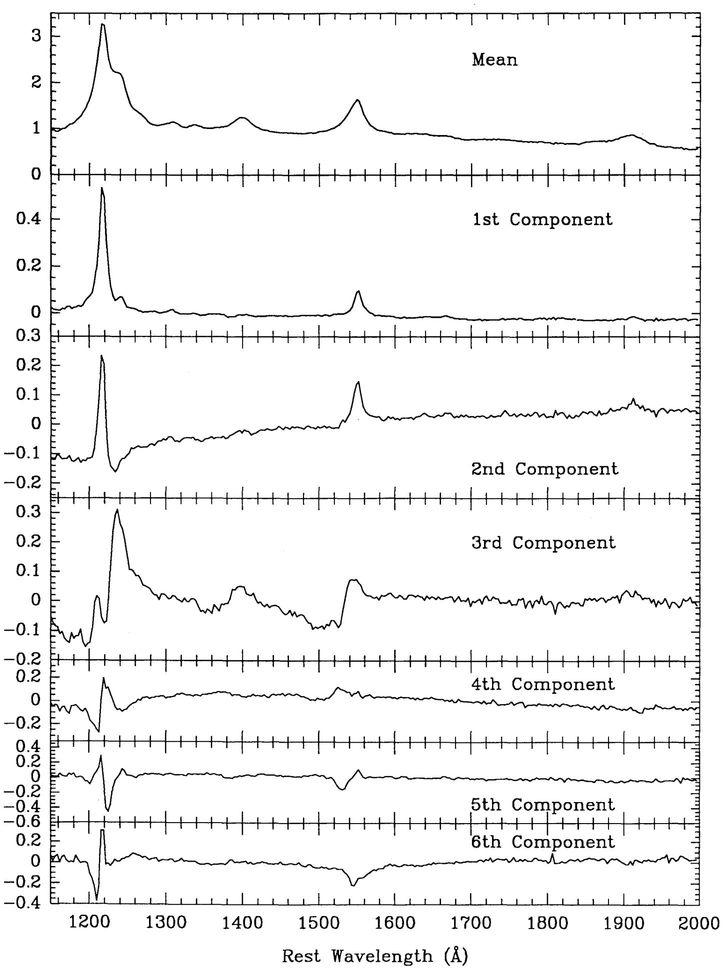

FIG. 1.—Mean spectrum and the first six principal components. All fluxes are $F(\lambda)$. Approximate errors for the components, expressed as standard deviations, are 0.003 for the first component, 0.006 for the second, 0.010 for the third, and 0.012 for the fourth, fifth, and sixth components.

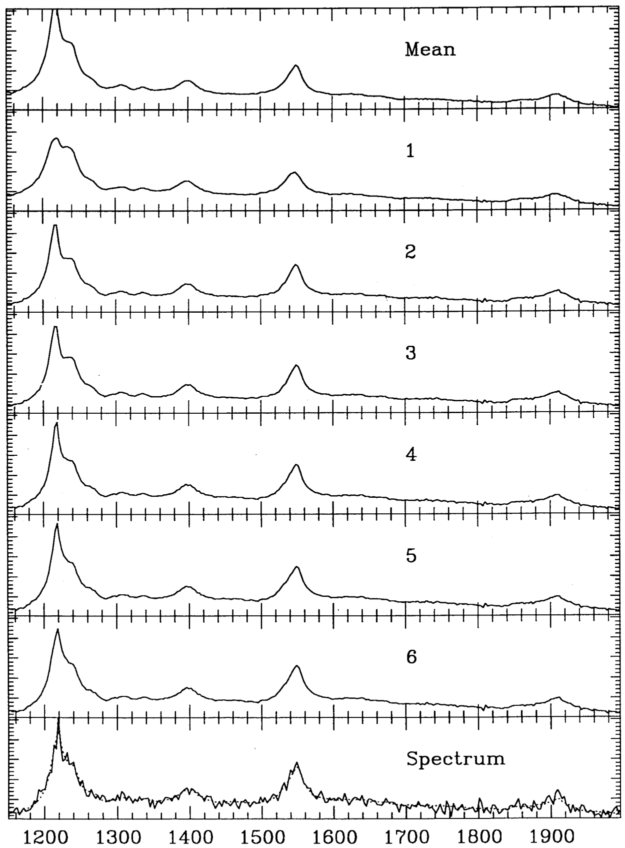

FIG. 2*a*

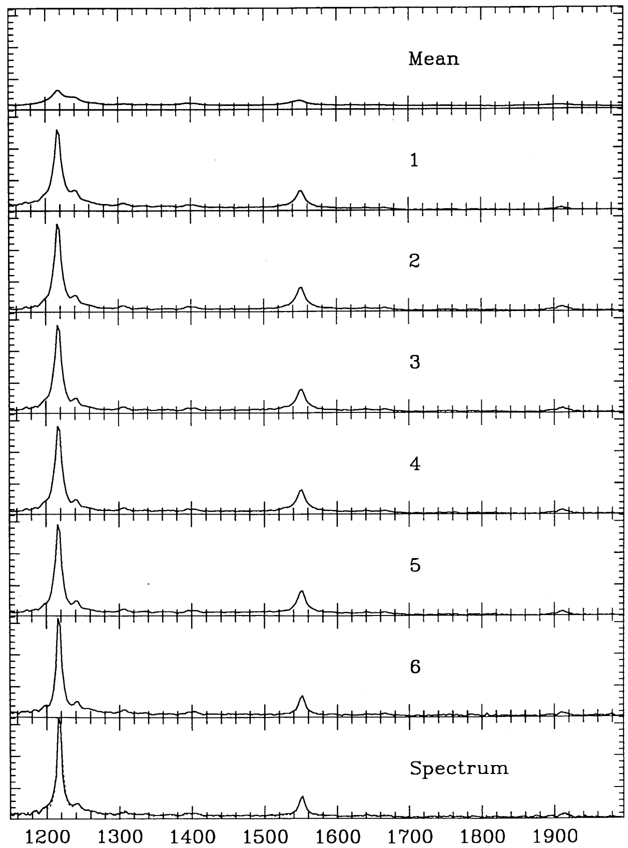

FIG. 2*b*

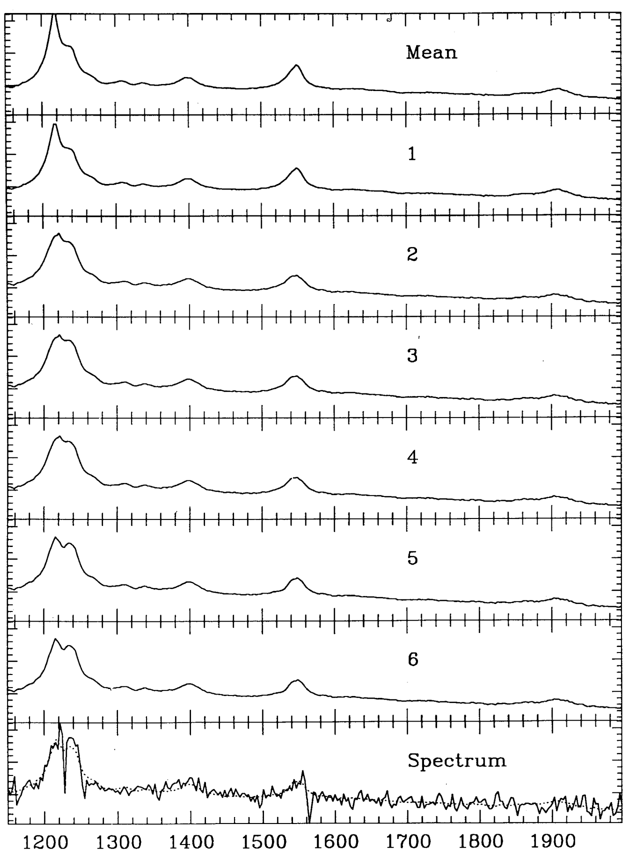

FIG. 2*c*

FIG. 2.—Reconstructions of three QSO spectra: (*a*) a randomly chosen QSO, (*b*) the QSO with the largest rms deviation from the mean spectrum, (*c*) the QSO with the largest rms deviation from a six-component reconstruction. The top panel of each figure shows the mean spectrum, and the next six panels show 1–6 component reconstructions of the spectrum. In the bottom panel the actual spectrum is shown (*solid line*) along with the six-component reconstruction (*dotted line*).

3.6. *Quantitative Assessment of the PCA*

Any QSO spectrum in the sample can be reconstructed by adding or subtracting suitably weighted principal components to the mean spectrum. Figures 2*a*–2*c* illustrate the component-by-component reconstruction of three QSO spectra. Figure 2*a* shows a typical spectrum; Figure 2*b*, the spectrum with the largest rms departure from the mean; and Figure 2*c*, the spectrum with the largest residual from a six component reconstruction. In the top panel of each, the mean spectrum is plotted. In the next six panels, the appropriately weighted component is added to the previous sum and plotted. Finally, the bottom panel presents the actual spectrum (solid line) along with the six-component reconstruction. The figures demonstrate how rapidly the reconstructions approach the observed spectra. In all cases, the correspondence between the spectrum and the four-component reconstruction is remarkably good.

Although Figure 2 provides a qualitative indication of how few principal components are required to achieve a satisfactory reconstruction of an arbitrary QSO spectrum, some quantitative measure of this convergence is required. The mismatch between a spectrum and its reconstruction arises from three sources: (1) intrinsic variations in the QSO spectra which have not been accounted for by the components employed, (2) differences in the spectral properties arising from factors unrelated to QSOs, notably absorption lines caused by intevening systems, and (3) random fluctuations resulting from the limited S/N of the observations.

Two distinct methods have been employed to assess the success of the PCA: the χ^2 statistical and the residual intrinsic variance.

3.6.1. χ^2 *Assessment*

The number of principal components needed to reconstruct the survey spectra to a level at which differences remaining between the survey spectra and the reconstructions can be accounted for by noise can be determined using the χ^2 statistic. The S/N of each spectrum is known for all wavelength bins, and a measure of the match of the reconstruction to the spectrum can be calculated by reconstructing all the spectra using a particular number of components and comparing the differences between the reconstructions and the spectra with the spectrum error arrays. Ideally this process would be repeated with increasing numbers of components, until the distribution of χ^2 values was consistent with noise being the only cause of spectrum-reconstruction differences.

Even using 20 components, the χ^2 statistic indicates that a substantial minority of QSOs are not reconstructed successfully. On inspection, strong intervening absorption systems were found in all such spectra. Since the absorption systems occur at essentially random wavelengths, they cannot be adequately reproduced using only a small number of components.

An attempt was made to exclude the intervening absorption systems from the χ^2 measurement, so as to determine the number of components necessary to reconstruct the features of the survey spectra excluding absorption lines. Lyman-α forest absorption lines were excluded by computing the χ^2 values only at wavelengths longward of the peak of the Lyman-α emission line. The effect of intervening absorption at random wavelengths was removed by excluding wavelength regions showing greater than 3 σ negative deviations from the reconstructions. BAL QSOs were not used in the analysis.

The distribution of the significances of the χ^2 statistic between the spectra and 15 component reconstructions is shown in Figure 3. If the differences were the result of noise, the distribution should be uniform, as it is over most of the plot. The peak at high probabilities shows that $\sim$5% of the spectra are being over-fitted, that is, the spectra can be adequately reconstructed using fewer than 15 components. The peak at low probabilities results from the $\sim$15% of spectra still not adequately reconstructed. The spectra which required a large number of components in their reconstruction were inspected visually, and all showed intervening absorption systems which had not been completely removed using the simple absorption-line identification procedure.

Excluding the subset of QSOs with BAL features the intrinsic continuum and emission-line spectra of all QSOs in the LBQS subsample can be reconstructed with fewer than 16 components.

3.6.2. *Residual Intrinsic Variance Assessment*

While the χ^2 statistic provides a measure of the number of components necessary to reconstruct QSO spectra, it is less suitable for comparing the importance of different components, as the χ^2 measure of the difference between a spectrum and a reconstruction depends strongly on the S/N of the spectrum. In this section the relative contributions of the components to the reconstructions are measured using the residual intrinsic variance, which is not strongly dependent on the S/N.

The first step in the calculation of the residual intrinsic variance is to measure the difference between each spectrum and its reconstruction by summing the squares of the differences of all the wavelength bins. This residual variance comprises intrinsic differences between the reconstructions and the spectra and noise. The contribution expected from noise is calculated by summing the squares of the estimated flux errors associated with the wavelength bins. As the intrinsic variation of QSO spectra should be uncorrelated with the noise, the residual intrinsic variance is the difference between the total residual variance and the noise variance.

Figure 4 shows the fraction of the residual intrinsic variance accounted for by each of the first 10 principal components. The first component alone reduces the residual variance by over a half, but the reduction produced by each additional component diminishes rapidly, and $\sim$10% of the intrinsic residual variance is not accounted for even with 10 components. The distribution of intrinsic residual variances from 10 component reconstructions shows a tail, corresponding to spectra which are not adequately reproduced. Once again, visual inspection

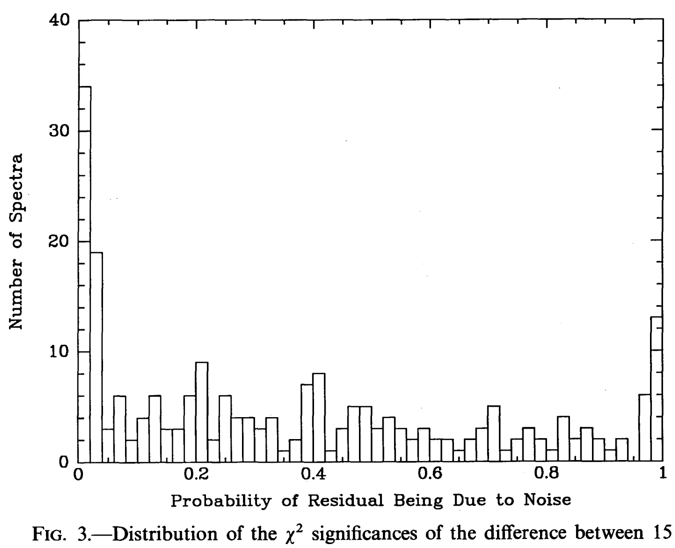

FIG. 3.—Distribution of the χ^2 significances of the difference between 15 component reconstructions and observed spectra.

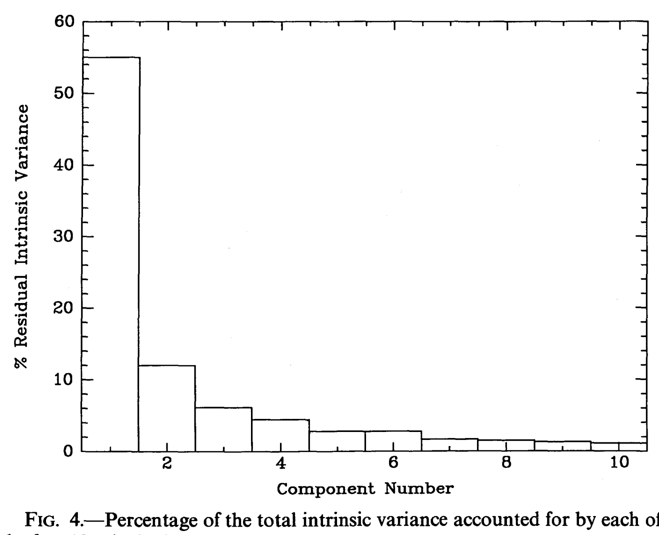

FIG. 4.—Percentage of the total intrinsic variance accounted for by each of the first 10 principal components.

of the spectra with large residual intrinsic variance showed the presence of strong intervening absorption systems.

The first few principal components thus account for the bulk of the intrinsic diversity of QSO spectra, with the first three components accounting for ~75% of the intrinsic sample variance (Fig. 4).

3.7. *Nonlinear Correlations*

Applying PCA directly to the QSO spectra has the benefits of simplicity and model independence, but it may exaggerate the apparent complexity of the spectral sample. This occurs when a change in a physical parameter produces a nonlinear change in the fluxes as a function of wavelength. Stellar spectral classification provides an illustration. The continua of stars are characterized by Planck functions of different temperature, and as the temperature changes, the peak of the flux distribution migrates through the spectrum, causing the flux to increase in some wavelength bins and decrease in others. It is not possible to reproduce this behavior using a mean spectrum and a single linear combination of fluxes as a function of wavelength, that is, with one principal component.

For QSO spectra, a similar example is provided by the change in the width of emission lines which results in a redistribution of flux in the vicinity of the line. The wavelength range affected is a function of the size of the width change, and the effects of the flux redistribution cannot be reproduced using a single linear combination of fluxes as a function of wavelength. Consequently, PCA has to generate a series of Fourier-like components, which combined with different weights can reproduce emission lines of all widths. Components 4–6 in Figure 1 are examples.

4. RESULTS AND DISCUSSION

In this section, the first three principal components are discussed in detail and then used as the basis of a classification scheme. These principal components demonstrate the *range* of properties exhibited by QSO spectra, not the average properties, which must be obtained from the mean spectrum (the first panel in Fig. 1).

4.1. *The First Component: Emission-Line Cores*

Strong modulations at the wavelengths of Lyman-α and C IV dominated the first principal component (see Fig. 1), which by definition is the largest linear source of diversity in the population of survey spectra. The velocity widths of the modulations are substantially narrower than the widths of the emission lines in the mean spectrum suggesting that much of the variation in emission-line equivalent widths among QSOs results from the changing contribution of "line cores." As a consequence, there is an anticorrelation between emission-line equivalent width and emission-line velocity width in the sense that emission lines with large equivalent widths tend to be narrow and have large peak fluxes. Although it is not evident in Figure 1, the first component does have a hard continuum relative to the mean spectrum. Therefore, if this component is added to the mean spectrum with a positive weight, the spectrum will be hardened. The first component thus contributes to the diversity of the sample spectra both by changing the line core strengths and the continuum slope, the former effect having a ninefold greater contribution to the total sample variance than the latter. The relatively small continuum contribution is significant however, since the component weights are large relative to those of subsequent components.

As a check of the conclusion that the line cores demonstrate the most variability, we have quantified the relative contributions of the Lyman-α and C IV line cores and line wings to the diversity of the spectra in the sample by measuring the variance of the line fluxes as a function of velocity from line center (defined relative to the redshift of the QSO), using the composite spectrum from Francis et al. (1991). The standard deviation of the line flux in the central 2000 km s^{-1} was ~70% of the mean flux, while that in the wings was only ~30%.

The relative strengths of the features in the first component are different from those of the corresponding emission lines in the mean spectrum. Thus the ratios of different emission-line fluxes are proportional to the line equivalent widths. This is especially evident for the Lyman-α to N V ratio, which is much greater than in the first component than in the mean spectrum.

The separation in wavelength between the centroids of C IV and Lyman-α is smaller in the mean spectrum than in the first principal component (Fig. 5). Thus, in objects with weak lines (small first-component weights), C IV is blueshifted relative to Lyman-α. As the line strengths in the QSO spectra change, so does the ratio of Lyman-α to N V, and hence the centroid of the blend shifts, raising the possibility of a systematic change in the measured redshifts of QSOs with different line strengths. The

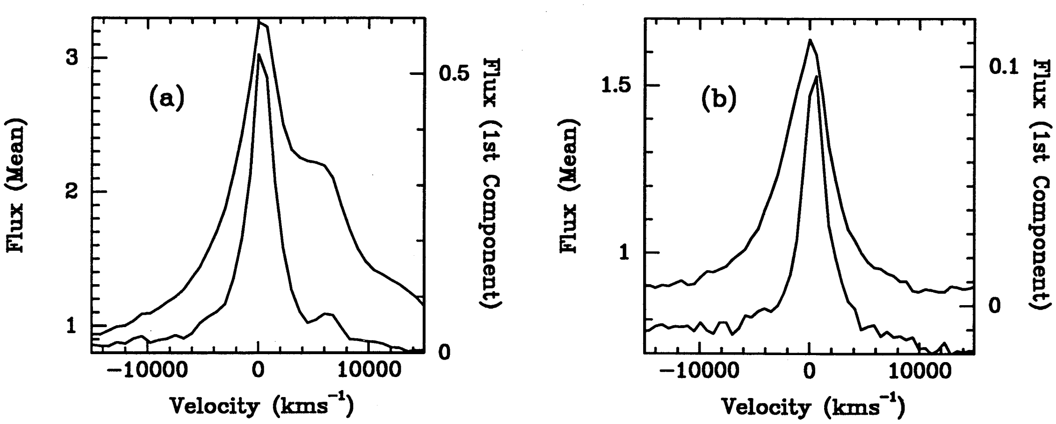

FIG. 5.—Comparison of line profiles in the mean spectrum and first principal component. Panel (*a*) shows Lyman-α + N V, panel (*b*) shows C IV. In each panel the upper profile is that of the mean spectrum; the lower profile, the first component. The profiles are plotted against velocity with respect to the nominal wavelengths of the transitions, 1216 and 1549 Å, respectively.

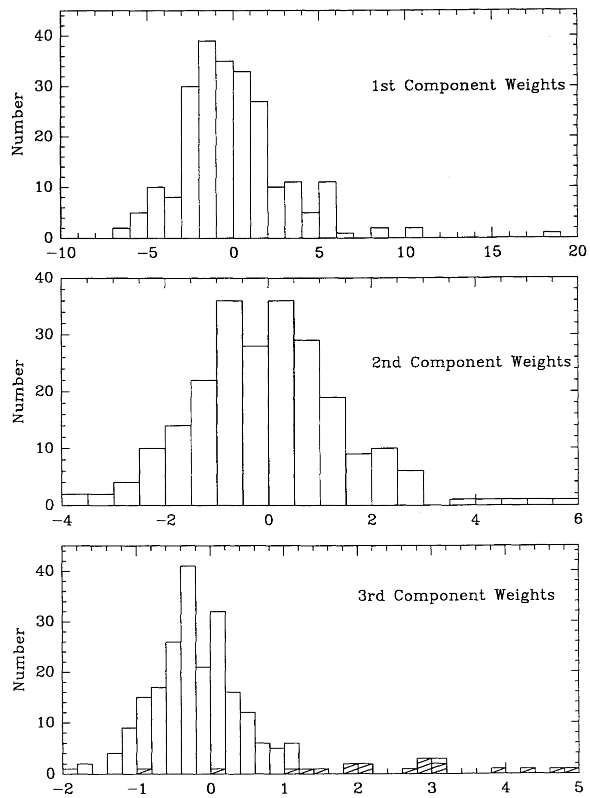

FIG. 6.—Distributions of the weights of the first three principal components. Strong BAL QSOs are shaded in the third-component histogram—the two BAL QSOs with the smallest third-component weights are unusual with extremely red continuum slopes and multiple BAL troughs.

magnitude of such an effect was investigated by reconstructing spectra with different first-component weights and measuring their redshifts by cross-correlation with the LBQS composite spectrum (Francis et al. 1991). Strong-lined objects were assigned higher redshifts than those with weak lines, but the magnitude of the effect ($\Delta z \sim 0.0005$) could account for only 20% of the observed line shifts.

The histogram of the weights of the first component (see Fig. 6) shows a non-Gaussian distribution. The extended tail to high component weights is made up of QSOs with strong, narrow emission lines which match closely the properties of the "narrow-line QSO" class identified by Baldwin et al. (1988).

These characteristics of the first, or "line-core," component mean that as its weight increases, the equivalent widths of the high-ionization emission lines increase; their full width at half-maximum (FWHM) decreases; their profiles become more sharply peaked; the rest wavelength separation of C IV and Lyman-α increases; and the line ratios change in the sense that Lyman-α becomes stronger relative to C IV, while N V, Si IV + O IV], and Al III become weaker. The trends in line profiles are evident in the C IV profiles illustrated in Figure 7.

The above progression could plausibly be explained either by (1) single-zone BLR models where a monotonic change in the strength of the first component results simply from a monotonic change in the angle from which the QSO is viewed, or (2) two-zoned BLR models with distinct line-wing and line-core regions.

As an example of the single-zone model, Wills, Fang, & Brotherton (1991) proposed that the anticorrelation between the FWHM and equivalent width of C IV might be explained if QSOs consist of an accretion-disk continuum source and a biconical outflow of line-emitted clouds with an axis perpendicular to the plane of the disk. Gas velocities in this biconical

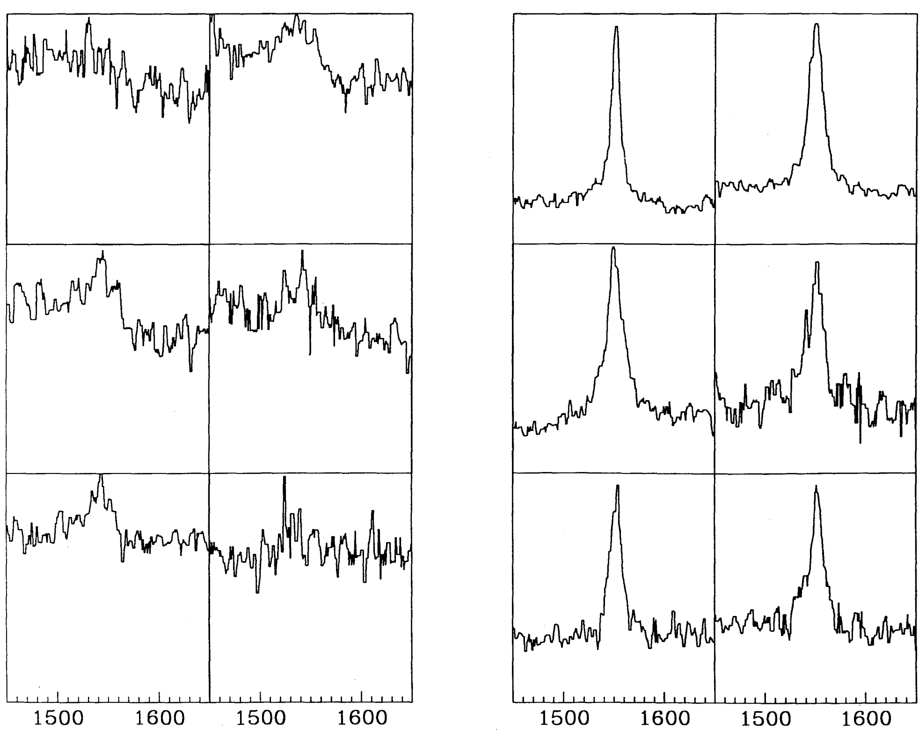

FIG. 7.—C IV profiles from spectra taken from the extremes of the distribution of first-component weights. QSOs with strong associated absorption or broad absorption affecting the C IV profile have been excluded. Profiles are plotted in relative flux with zero at the bottom of each panel.

outflow are greater along the outflow axis than perpendicular to it. When the QSO is observed down the flow axis, the lines are broad, because of the higher velocities along the line of sight, and the equivalent widths are low because of the reduced line contrast against the strong continuum. While such a model accounts for this particular behavior of the C IV emission line, it cannot explain the near-constant velocity width and equivalent width of the emission-line wings, because when viewed from perpendicular to the cone axis, the line wings would show lower velocity widths than when viewed down the axis.

Relative line shifts could be produced in such a model if the optically thick portion of the accretion disk were comparable in size to the line-emitting region, and if different emission lines were preferentially radiated at different distances from the continuum source (as suggested for low-luminosity objects by variability measurements [Clavel 1991]). In particular, if Lyman-α were radiated further from the continuum source than C IV, and the clouds were outflowing, then C IV would be blueshifted relative to Lyman-α when the QSO is viewed face on, due to the obscuration of the retreating flow by disk. In this case, the C IV–Lyman-α separation would be greatest when the equivalent widths were small, as observed here.

A correlation that is more difficult to explain is that between line equivalent width and line ratios. For the orientation model, the nature of the first component requires that N V, Si IV + O IV], and Al III be strongest when observed in the plane of the disk, while Lyman-α is strongest when observed perpendicular to the disk.

A two-zone line-emitting model, consisting of a line-wing zone characterized by velocities of 5000 km s^{-1} FWHM, and a line-core zone with velocities $\sim$2500 km s^{-1} FWHM is consistent with a number of the observed correlations. Note that the line-core region is not the traditional narrow-line region (NLR). All the line cores are resolved in the LBQS spectra, and they greatly exceed typical NLR widths. support for a two-zone model comes from the variability measurements of NGC 5548 and NGC 4151 (Clavel 1991), in which the line cores (defined as the central 3000 km s^{-1}) lag the continuum with roughly twice the delay of the line wings.

Plausible explanations for the range and distributions of line-core strengths in a two-zone model include changes in the volume emissivity of the line-core-emitting zone, continuum variability, and microlensing.

A change in the volume emissivity of the line-core-emitting clouds (or in the case of optically thick clouds, the covering factor of the cloud ensemble) could explain the dispersion in the strength of the first component. However, a successful model must explain why the line-core covering factor should vary more than that of the line wings.

If the zone in which the line wings are formed lies close to the continuum source, then its photoionized line emission will track variations in the continuum emission with a short lag. If the continuum variability time scale is comparable to the light travel time between the continuum source and the line-core region, then as the continuum flux varies, a range of line-core equivalent widths will be observed. However, observed continuum variability amplitudes (e.g., Hawkins 1986; Hook et al. 1991) of $\sim$0.2 mag are too small to account for the observed factor of $\sim$3 equivalent width range. Furthermore, in order to explain the asymmetric line-core strength distribution, QSO flux variations would have to be characterized by long periods of slowly increasing luminosity punctuated by sudden declines.

The continuum source and line-wing zone of some QSOs could be microlensed (Canizares 1982). Objects of stellar mass have an Einstein radius of comparable angular size to the inner BLR of a QSO. QSOs with low equivalent widths would be those with a lensing mass along the line of sight amplifying the continuum and line wings. Because of the steepness of the QSO luminosity function, strongly amplified QSOs would dominate the sample, leading to the observed asymmetric line-core equivalent-width histogram. However, as Canizares points out, a nearly critical density of lensing objects is required for this model to be viable.

4.2. *The Second Component: Continuum Slope*

The second component is dominated by a large-scale continuum slope. Therefore, the dominant effect of a change in this components weight will be a change in the hardness of the QSO continuum. There are also modulations at the positions of Lyman-α, N V, C IV, and Al III + C III], indicating that emission-line strengths are correlated with the continuum slope. For Lyman-α and C IV, the lines are stronger when the continuum is soft, while for N V, the converse is true. The amplitude of the "N V modulation" is, however, dependent upon the shape of the underlying continuum across the Lyman-α + N V region. For the Lyman-α and C IV emission lines, the correlation with the large-scale slope is in the opposite sense to that seen in the first component. While the large-scale continuum slope dominates this component, the reverse is not true, and this component does not alone dominate the continuum slope. The continuum slope in this component is roughly 3 times greater than that in the first component, but the first component weights show a factor of 3 greater range. Therefore the contributions of the first and second components to the variation in continuum slopes are comparable.

In what follows, the continuum is parameterized as a power law joining the rest-frame wavelength regions 1320 and 1800 Å, following the convention $F_\nu \propto \nu^\alpha$. While it is clear from the LBQS composite spectrum (Francis et al. 1991) that the 1800 Å window is contaminated by Fe II emission, the smoothness of the first and second components suggests that Fe II emission is not a strong constituent of these components. This gives confidence that the large-scale shape of the second component reflects a variation in some continuum process, either due to the presence of dust or to a real variation in the emitted continuum. The presence of emission features at familiar wavelengths in the second components suggests that the continuum slope correlates with the strengths of some of the emission lines. In particular, the strength of Al III + C III] correlates strongly with α, those of Lyman-α and C IV also correlate with α but show a large scatter, while N V and Si IV + O IV] show little variation. A successful model for the QSO phenomenon must explain the range in continuum slopes and the line-continuum correlations.

Intervening dust absorption has often been invoked to explain the observed range of QSO continuum slopes (see, e.g., Cheng, Gaskell, & Koratkar 1991). If the hardest QSO spectra in the LBQS are unabsorbed, if their continuum slopes are representative of the unobscured parent population, and if the whole range of continuum slopes results from obscuration of some canonical underlying continuum, then the QSOs with the softest spectra are absorbed by *over* 2.5 mag at 1320 Å in the QSO rest frame, assuming the dust lies at the QSO redshift and follows Galactic, LMC, or SMC extinction relations (Fall &

Pei 1989). This monochromatic absorption corresponds to an extinction $E(B-V)$ of ~ 0.2 mag. For a Galactic dust-to-gas ratio, a neutral hydrogen column density N_H of order 10^{21} cm^{-2} is implied, which, as noted by McDowell et al. (1989), would strongly attenuate the soft X-ray emission and would also introduce a substantial 2200 Å feature were the dust extinction law like that of LMC or Galactic dust. Finally, if the dust were external to the BLR, it would obscure the line and continuum emission equally and hence leave the equivalent widths unchanged. Thus to produce a correlation of line equivalent widths with the continuum slope requires that some of the dust lie within the BLR. If dust were added to the BLR clouds, it might be expected to preferentially absorb the more optically thick lines such as Lyman-α, leaving intercombination lines such as C III] relatively unaffected. This is opposite to the behavior observed—C III] is strong relative to Lyman-α in the QSOs with the hardest spectra. On this basis, a variable amount of dust is unlikely to be responsible for the observed range of the second component. This does not imply that heavily dust-shrouded QSOs do not exist, since because of the steepness of the QSO luminosity function, an optical magnitude-limited sample such as the LBQS will be heavily biased toward the least absorbed QSOs.

A more likely explanation for the continuum variation implied by the range of second-component weights is a real dispersion in the emitted continuum. The observations reported here were all obtained in a rest-frame wavelength region dominated by the blue bump UV excess (Malkan & Sargent 1982). If the measured continuum slopes are extrapolated to 1 μm, they predict a range of UV-IR flux ratios similar to that presented by McDowell et al. (1989). Therefore, the range of continuum slopes measured for the LBQS QSOs may reflect the strength of the UV bump with the reddest LBQS spectra corresponding to the weak-bump QSOs identified by McDowell et al.

Can the correlations of continuum slope with equivalent width be explained by photoionization? Both the intensity and shape of the ionizing continuum will vary with the strength of the blue bump. As the observed continuum becomes harder, the ratio of continuum flux at the Lyman limit to continuum flux in the near-UV ($\lambda\lambda$1150–2000) will increase, the number of recombinations per observed continuum photon will rise, and hence the equivalent width of recombination lines such as Lyman-α (other things being equal and provided the emitting gas is optically thick to the continuum) will rise, in agreement with the sense of the weak observed correlation. However, the correlations between the continuum slope and the other emission-line equivalent widths obey no clear pattern: C IV behaves like Lyman-α, while the lower ionization Al III + C III] correlates more tightly with the continuum hardness, and N V and Si IV + O IV] show little variation.

4.3. *The Third Component: Broad Absorption Lines*

The third component (see Fig. 1) exhibits correlated modulations over a significant wavelength range to the blue of the C IV, Si IV + O IV], and N V emission lines. In addition to these broad features there are positive modulations at the wavelength of N V and, to a lesser degree, C IV and Al III + C III], showing that the presence of broad features blueward of the high ionization lines is anticorrelated with the emission-line strength.

The distribution of the weights of this component (Fig. 6) shows that while the majority of QSO spectra have third-component weights near zero, a minority have large negative values—that is, the broad features are seen in absorption. There is an excellent correlation between objects with large negative third-component weights and QSOs which would be classified as BAL QSOs. The third component thus identifies BAL QSOs. Two potentially important characteristics of BAL QSOs are substantiated by the nature of the third component and the distribution of its weights among the QSOs in the sample.

First, if BAL QSOs were a totally different class of objects from normal QSOs, one would expect the histogram of third-component strengths to exhibit a bimodal form. That the observed histogram appears continuous suggests, but does not require, that BAL QSOs belong to the same population as normal QSOs, differing only in the magnitude of some physical parameter. If the parameter is viewing angle, this corresponds to the "single-population" model of Weymann et al. (1991), where all optically selected QSOs are intrinsic BAL QSOs, but are seen as such only when one's line of sight pierces a BAL cloud. The continuity of the distribution of third-component weights suggests further that a class of hitherto unrecognized weak-BAL QSOs exists.

Second, the presence of apparent emission features in the third component suggests that BAL QSOs have enhanced N V emission and a smaller enhancement of Al III + C III] and C IV emission. An attempt was made to evaluate the significance of these enhancements in the following way: The largest source of uncertainty in the equivalent-width excesses comes from the small numbers of BAL QSOs observed. To evaluate this, the standard deviations of the equivalent widths of the three emission features in non-BAL QSOs were computed using two-component reconstructions (see § 4.4). These values were divided by the square root of the number of BAL QSOs to give an estimate of the typical difference to be expected between BAL and non-BAL QSOs. For N V, the flux at 1240 Å was used in place of an equivalent width, due to the problems of deblending from Lyman-α. The excesses of both N V and Al III + C III] observed in the third principal component were found to be significant at better than the 99% level, but that of C IV was only a 1 σ effect, and hence cannot be regarded as significant. These trends are the same as those elucidated by Weymann et al.

4.4. *A Quantitative Classification Scheme*

Since all LBQS QSO spectra can be adequately reconstructed by adding the first three principal components, suitably weighted, to the mean spectrum, the weights of these three components can be used to classify QSO spectra. Sample LBQS spectra classified by the weights of their first two components are presented in Figure 8. The spectra are located within the plot such that those with positive first-component weights are at the right, those with positive second-component weights are at the top. Therefore, line cores are strongest to the right, and the continuum is hardest at the bottom of the plot. Properties such as line profiles, line ratios, equivalent widths, and continuum slopes all vary smoothly across the plot.

A spectrum can be classified within this scheme in one of two ways. First, some fitting algorithm can be used to find the combination of component weights which best reconstructs the spectrum. We are prepared to make digital copies of the components available for this purpose. Second, such parameters as line equivalent widths and continuum slopes can be used to locate a spectrum within this classification scheme,

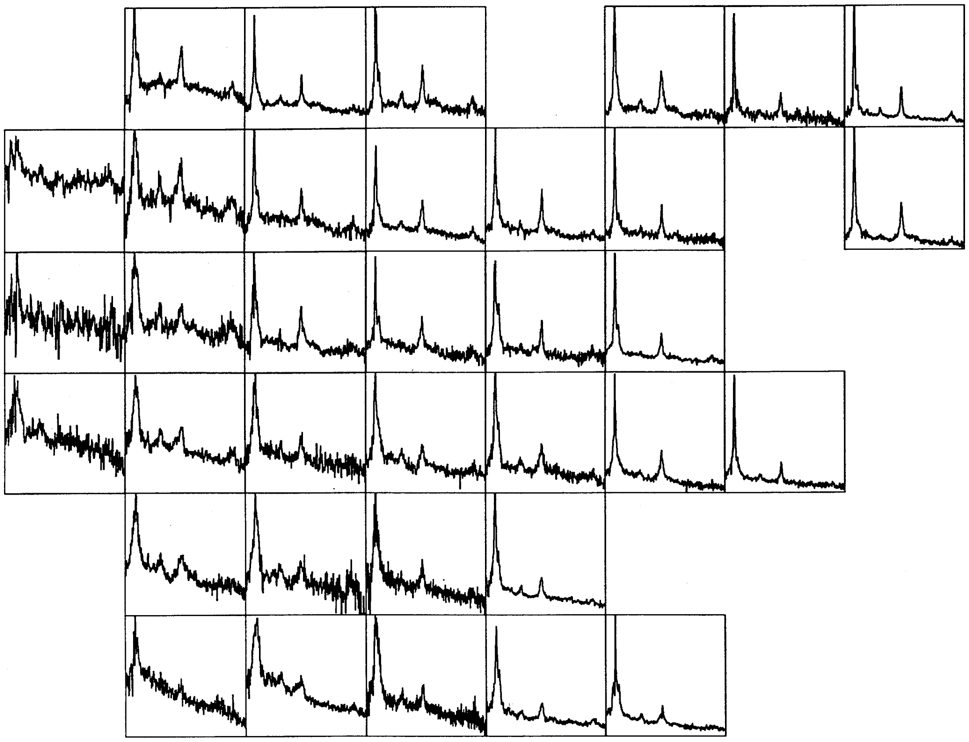

FIG. 8.—LBQS spectra, chosen at random, plotted by their two-component classification. The first-component weight of the spectra increases to the right, ranging from −5 to 11, while the second-component weight increases upwards, ranging from −3 to 3. Thus the line-core equivalent widths increase to the right, while the continuum is softest at the top. All spectra show the rest-frame wavelength range 1150–2000 Å, with relative fluxes per unit wavelength. BAL QSOs have been excluded from the plot. If no LBQS spectrum has a two-component classification in a given range, the corresponding part of the plot is left blank.

using the relations between the principal component weights and these parameters derived below and shown in Table 1. The LBQS spectra are classified within the scheme as a byproduct of the definition of the principal components, which is equivalent to choosing the component weights to minimize the root-mean-squared residuals between the spectra and their reconstructions.

In order to facilitate comparison with existing work, linear relationships between the three-dimensional classification scheme (i.e., the strength of the first three principal components) and such familiar parameters as continuum power-law indices and emission-line equivalent widths have been derived using the following procedure: First, the zero point of the relationship was established by measuring the parameters from the mean spectrum. Then "synthetic" spectra were constructed with unit positive and negative weights of each of the three components in turn, and the same parameters were measured from these spectra. (Measurements of spectra with positive and negative weights were carried out as a consistency check.) The partial derivative of each parameter with respect to each principal component weight was thus determined. Using these coefficients, the value of the appropriate parameter can then be determined for any observed spectrum from the weights of its first three principal components. The efficacy of this procedure relies on the existence of linear relationships between the parameters and the component weights. This is a good approximation except in the case of the line widths (see § 3.7).

Emission-line equivalent widths were measured from the mean and synthetic spectra by connecting selected wavelength regions on either side of the emission feature with a straight line, and integrating the flux above this line. The wavelength limits for the integrations were the following: $\lambda\lambda$1180–1280 Lyman-α + N V; $\lambda\lambda$1290–1325 O I; $\lambda\lambda$1350–1450 Si IV + O IV]; $\lambda\lambda$1440–1610 C IV; and $\lambda\lambda$1830–1970 Al III + C III]. As noted above, the continuum power-law slope was measured using continuum windows at 1320 and 1800 Å. The former has a contribution from C II and Si IV + O IV], the latter from Fe II (Francis et al. 1991). A velocity width for C IV was measured as described in Robertson (1986). The continuum used in the measurement of the third (BAL) component properties is shown in Figure 9. The coefficients for the various relationships are given in Table 1.

TABLE 1

CONVERSION OF PRINCIPAL COMPONENT WEIGHTS TO CONVENTIONAL PARAMETERS

Feature	Component Coefficients				Units
	Constant	First	Second	Third	
EW(Lyman-α)[a]	57	7.7	7.3	6.2	Å
EW(Si IV/O IV] BAL)	...	...	...	2.45	Å
EW(Si IV/O IV])[b]	8.8	0.2	0.6	0.5	Å
EW(C IV BAL)	...	...	...	8.5	Å
EW(C IV)	26	2.6	2.5	1.1	Å
EW(Al III/C III])[c]	21	1.3	0.15	3.4	Å
FWHM(C IV)	6600	−60	−500	...	km s^{-1}
α[d]	−0.7	0.1	−0.3	...	

[a] Lyman-α forest absorption affects both the measured emission and the continuum values used.

[b] In the first component, Si IV/O IV] is very weak and shifted to the red.

[c] In the first component, no Al III emission is seen—only emission at or around 1909 Å contributes.

[d] α is defined such that $F(\nu) \propto \nu^{\alpha}$.

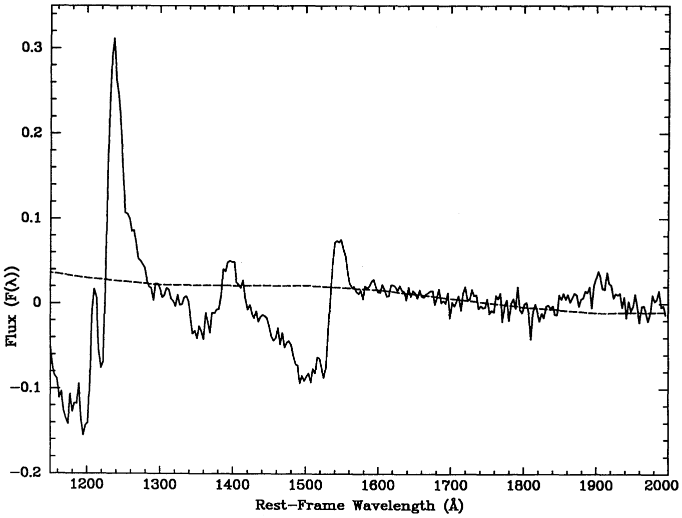

FIG. 9.—The third principal component (*solid line*) and the continuum (*dashed line*) adopted for equivalent-width measurements of the emission lines.

Combined with the distribution functions of the principal component weights (Fig. 6), the relations in Table 1 provide a quantitative description of any of the LBQS spectra and elucidate the range of parameters observed in the sample as a whole. This simple description of the LBQS properties can be used, among other applications, to model the selection effects of the survey, to compare the LBQS to detailed emission-line models, and to compare the distribution of parameters to other surveys of QSOs.

Using the relations in Table 1, and the component weights of the LBQS spectra, the distributions of line strengths and continuum slopes can be computed. Figure 10 shows the distribution of C IV rest-frame equivalent widths, compared with two distributions taken from the literature (Wilkes 1986; Baldwin, Wampler, & Gaskell 1989). All three distributions have comparable mean and modal values, but the Wilkes sample has a larger proportion of strong-lined objects, perhaps reflecting differences in the QSO selection between the samples.

In Figure 11 the continuum slopes measured from the PCA of the high-ionization line region are compared with those of the radio-quiet subsample studies by Baldwin et al. (1989). Baldwin et al. investigated a slightly different rest wavelength region and used continuum windows at 1450 and 2225 Å to determine their continuum slopes. The systematic change in the derived power-law slope caused by using different continuum windows was quantified by measuring the continuum slope of the LBQS composite spectrum using both pairs of continuum windows. The Baldwin windows produce a larger index (a harder slope) by 0.3, due to a combination of the continuum's becoming softer at shorter wavelengths and the presence of Fe II emission at 1800 Å. When the effect of the different continuum windows is taken into account, both the mean and the range of the power-law distributions are consistent with each other. Specifically, a KS test indicates that, after the 0.3 systematic correction is applied to the Baldwin sample, one can reject the null hypothesis that the two samples are drawn from the same parent population at only the 80% confidence level. After correction the mean index of the Baldwin sample is 0.91 and its standard deviation is 0.34, while the LBQS has a mean of 0.67 and a standard deviation of 0.5, in reasonable agreement with one another given the small size of the Baldwin sample.

Figure 12 illustrates the correlations between the continuum slope and the emission-line equivalent widths derived from the two-component reconstructions. Table 2 gives the results of regression-line fits to these scattergrams. In this analysis, BAL QSOs were excluded, and the errors in alpha were ignored. As discussed in § 4.2, C IV and Lyman-α correlate with continuum hardness but show a large scatter, while Al III + C III] correlates more tightly with the continuum hardness. The Si IV + O IV] correlation, although formally significant, should be viewed with caution given the weakness of the fea-

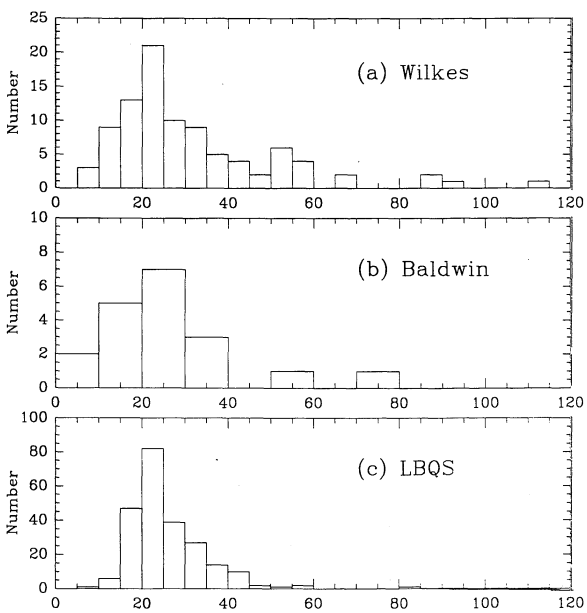

FIG. 10.—C IV rest-frame equivalent width distributions. Panel (*a*) is taken from Wilkes (1986), panel (*b*) is the optically selected subsample of Baldwin et al. (1989), and panel (*c*) shows values computed from the two-component LBQS classification.

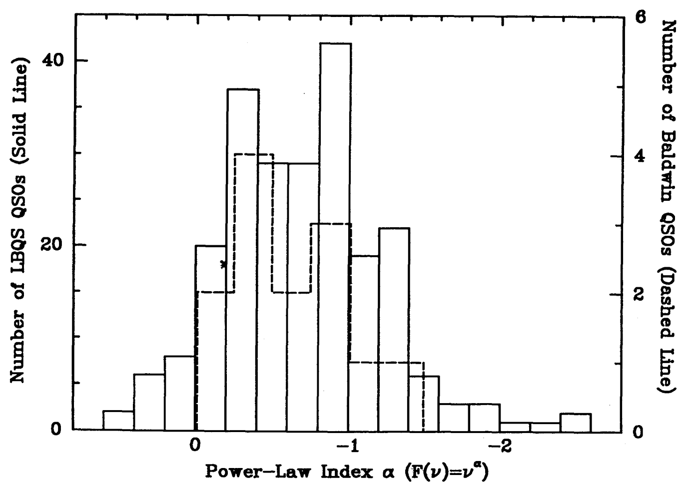

FIG. 11.—The distribution of continuum power-law indices for the LBQS (*solid line*) measured between 1320 and 1800 Å, and those from the optically selected subsample of Baldwin et al. (1989) (*dashed line*) measured between 1450 and 2225 Å.

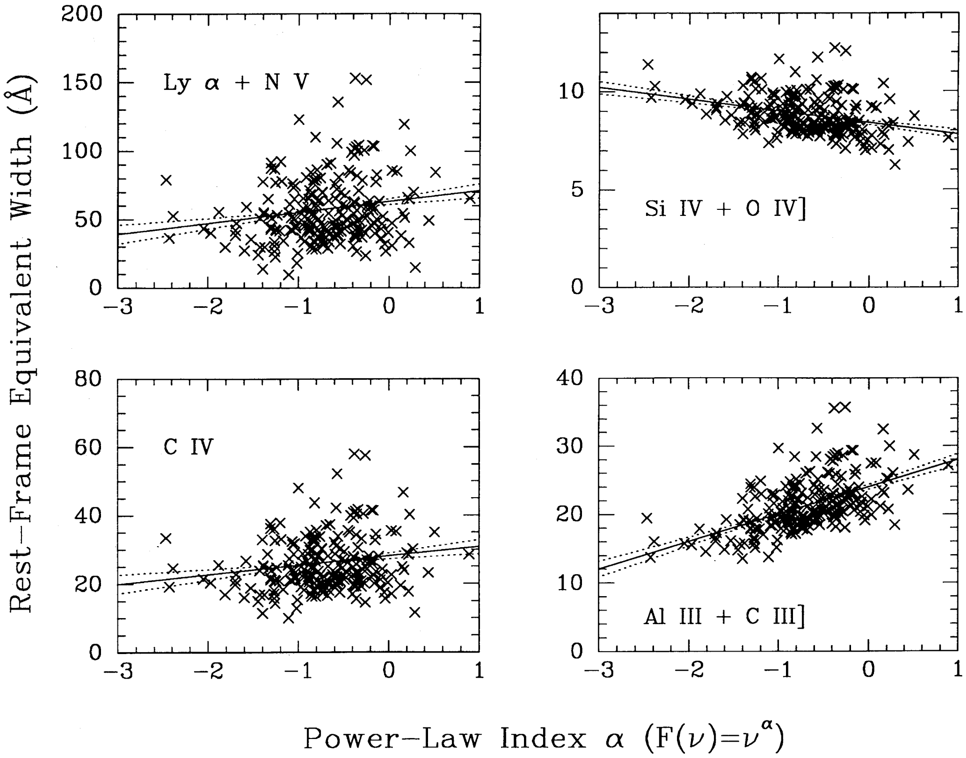

FIG. 12.—Rest-frame equivalent widths of four emission lines as a function of the continuum slope measured between 1320 and 1800 Å. All measurements are computed from the two-component reconstructions. Solid lines are the best-fit regression lines; dotted lines show the $\pm 1\,\sigma$ slope limits (Table 2).

tures and the resulting strong dependence of the equivalent width on the adopted continuum windows.

4.5. *The Baldwin Effect*

Recent analyses of samples of optically selected QSOs (Cristiani & Vio 1990; Zamorani et al. 1992) have shown that the Baldwin effect—the anticorrelation between C IV equivalent width and absolute magnitude—is weak with a substantial scatter in the line equivalent widths at fixed absolute magnitude. Measurement of the C IV equivalent width from the reconstructed spectra would seem to offer a substantial advantage, compared to using the original spectra, because the noise is reduced significantly. However, the use of an LBQS subsample to investigate the amplitude of the Baldwin effect has the disadvantage of covering only a limited redshift and apparent magnitude range ($17.0 < m_{B_J} < 18.7$). This means that the QSOs in the sample investigated here span the rather limited absolute magnitude range of $-28 \lesssim M_B \lesssim -26$. On the other hand, the number of QSOs in the subsample is large, the S/N of the spectra is relatively high, allowing emission-line equivalent widths to be measured to relatively high precision, and the range of emission-line equivalent widths, $0 \lesssim EW_{\rm C\,IV} \lesssim 100$ Å, is substantial.

TABLE 2

EMISSION-LINE EQUIVALENT WIDTH–CONTINUUM POWER-LAW INDEX CORRELATIONS

Emission Line	Constant	Slope
Lyman-α	63 ± 3	8 ± 3
Si IV + O IV]	8.4 ± 0.1	-0.6 ± 0.14
C IV	28 ± 1	2.7 ± 1.2
Al III + C III]	24 ± 0.4	4 ± 0.5

When the C IV rest equivalent widths, computed from the three-component reconstructed spectra, are plotted versus absolute magnitude, a weak Baldwin relation is observed. Specifically, a formal regression of the rest equivalent widht of C IV against the monochromatic absolute magnitude at 1450 Å, assuming a functional form of $EW_{\rm C\,IV} \propto L_{1450}^{\beta}$, gives $\beta = -0.14 \pm 0.05$. The details of this calculation are three: (1) The absolute magnitudes were calculated by integrating each spectrum over the IIIa-J passband to determine the scaling factor between the observed broad-band and monochromatic magnitudes. (2) BAL QSOs and objects showing strong associated absorption have been excluded from the analysis because of the ambiguities in establishing a continuum and determining "unabsorbed" emission-line equivalent widths. (3) The errors were estimated by assuming that the independent scatter at any fixed luminosity is equal to the standard deviation of the equivalent widths of the entire sample—a good approximation since the size of the Baldwin effect is a small fraction of the intrinsic scatter.

The value of β so derived is consistent with the values obtained by Zamorani et al. (1992) and by Kinney, Rivolo, & Koratkar (1990), but significantly smaller than the values of Baldwin et al. (1989) and Smetanka et al. (1991) (see Table 3).

However, by using the C IV equivalent width from three-component reconstructions to estimate the slope of the Baldwin effect, we tacitly assume that the line-wing equivalent widths are luminosity-independent. This is in keeping with the results of § 4, in which it was shown that the wings are responsible for only a small fraction of the total variation in the line equivalent widths. But, since over the limited lumi-

TABLE 3

BALDWIN EFFECT—$EW(\text{C IV}) \propto L_{1450}^{\beta}$

Measurement Source	β	σ_{β}
LBQS—Three Component	−0.14	0.05
LBQS—Direct	−0.26	0.07
LBQS—Line Core	−0.15	0.05
LBQS—Line Wings	−0.07	0.03
Baldwin et al. 1989	−0.39	0.10
Cristiani & Vio 1990	−0.083	0.006
Kinney et al. 1990	−0.17	0.04
Smetanka et al. 1991	−0.45	0.08
Zamorani et al. 1992	−0.13	0.03

nosity range covered by the LBQS sample, the expected magnitude of the Baldwin effect is much smaller than the range of equivalent widths observed at any fixed luminosity, systematic correlations of the flux in the wings with luminosity could strongly modify any correlation seen in spectra reconstructed from only the first three components. In order to investigate this possibility, equivalent widths were measured for C IV directly from the spectra, using the same integration limits and continuum windows as in the measurements of the reconstructed spectra. These direct measurements are shown in Figure 13, along with the best-fit regression with $\beta = -0.26 \pm 0.07$. This slope differs from the three-component value by 1.4 σ and is more similar to the Baldwin and Smetanka values.

While the difference between the slope calculated for the direct measurements and that calculated for the reconstructed spectra is not highly significant, its origin needs, nonetheless, to be investigated. This was done by measuring directly from the spectra the equivalent widths for the line cores (measured in the range 1540–1560 Å) and for the remainder of the line separately. The line-core equivalent widths correlate closely with those measured from the three-component reconstructions and show nearby the same slope for the Baldwin effect, $\beta_{\text{core}} = -0.15 \pm 0.04$. The line-wing equivalent widths have a mean similar to the line cores (~ 14 Å), but show a weaker, but nonzero, correlation with luminosity, $\beta_{\text{wings}} = -0.07 \pm 0.03$.

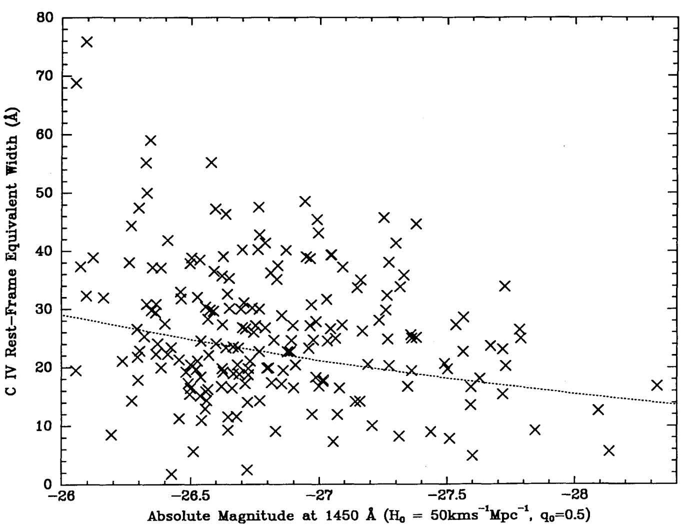

FIG. 13.—C IV equivalent widths as a function of the continuum absolute magnitude at 1450 Å. Equivalent-width measurements are not from the reconstructions but from the original spectra, as described in the text. The dashed line is a least-squares power-law fit to the data.

The combination of the two gives a slope close to that measured directly for the whole line. It is interesting to speculate that if this effect is real (the difference is only significant at the 1.6 σ level), then some of the difference in the published results may arise because of different choices for the continuum placement.

Table 3 summarizes the Baldwin effect slopes measured by the methods described above as well as those measured by a number of other recent studies.

Our search for the Baldwin effect using PCA-inferred C IV equivalent widths provides a cautionary note to the application of PCA to effects that may be rather subtle. Calculating the C IV equivalent width from only three-component reconstructions significantly diluted the inferred magnitude of the Baldwin effect, suggesting in this case that the reconstructions did not well represent the original spectra. Such reconstructions did, however, call attention to the question of whether the Baldwin effect differs between line cores and line wings, and closer examination revealed that, indeed, there may be such a difference. If this difference is found to be real, it may have important implications for models attempting to explain the physical basis of the effect.

5. SUMMARY

1. Principal component analysis, applied directly to QSO spectra, is a powerful tool for the parameterization, and subsequent classification, of rest-frame ultraviolet spectra observed at intermediate S/N and resolution.

2. Only three components are required to account for $\sim 75\%$ of the QSO sample variance contributed by the variation in the ultraviolet emission lines and continua. One component describes the emission-line core strength, the second the continuum slope, and the third the BAL features present in a subset of the spectra.

3. Broad-absorption-line QSOs are naturally identified as a subclass of the QSO population. The distribution of the third principal component weights supports a single-population model for BAL QSOs, wherein all optically selected QSOs may be incipient BAL QSOs. The BAL QSOs show excess N V and perhaps Al III + C III] emission relative to the QSO population as a whole.

4. The emission lines of the QSO sample show a spread of properties which can be arranged in a sequence ranging from broad, low-equivalent-width lines to narrow, high-equivalent-width lines. Other properties of the QSO spectra also vary along this sequence. For example, QSOs in the former category have higher N V/Lyman-α, N V/C IV, and C IV/Lyman-α line ratios than objects with lines characteristic of the latter category, and the wavelength separation between the Lyman-α and C IV emission lines is larger in spectra in the former category than in the latter.

5. The equivalent widths of the Lyman-α, C IV and Al III + C III] emission lines all correlate with the hardness of the continuum, the latter being the tightest correlation.

6. The LBQS sample demonstrates a Baldwin effect consistent with other studies. It is suggested that the correlation of the line-core equivalent widths with luminosity may be stronger than that of the line wings.

The LBQS is supported by National Science Foundation grants AST 90-01181 and 90-05117, for which we are grateful.

Computing support was provided by the Starlink Project which is funded by the UK SERC. P. J. F. was supported by an SERC postgraduate studentship. F. H. C. gratefully acknowledges the hospitality of the Institute of Astronomy and the support of a Royal Society Guest Research Fellowship. We heartily thank Ray Weymann for suggestions and criticisms.

REFERENCES

Baldwin, J. A. 1977, ApJ, 214, 679
Baldwin, J. A., McMahon, R., Hazard, C., & Williams, R. E. 1988, ApJ, 327, 103
Baldwin, J. A., Wampler, W. J., & Gaskell, C. M. 1989, ApJ, 338, 630
Boroson, T., & Green, R. 1992, ApJS, 80, 109
Canizares, C. R. 1982, ApJ, 263, 508
Chaffee, F. H., Foltz, C. B., Hewett, P. C., Francis, P. J., Weymann, S. L., Morris, S. L., Anderson, S. F., & MacAlpine, G. M. 1991, AJ, 102, 461
Cheng, F. H., Gaskell, C. M., & Koratkar, A. P. 1991, ApJ, 370, 487
Clavel, J. 1991, in Variability of Active Galaxies, ed. W. J. Duschl, S. J. Wagner, & M. Camenzind (Berlin: Springer), 31
Cristiani, S., & Vio, R. 1990, A&A, 227, 385
Fall, S. M., & Pei, Y. C. 1989, ApJ, 337, 7
Foltz, C. B., Chaffee, F. H., Hewett, P. C., MacAlpine, G. M., Turnshek, D. A., Weymann, R. J., & Anderson, S. F. 1987, AJ, 94, 1423
Foltz, C. B., Chaffee, F. H., Hewett, P. C., Weymann, R. J., Anderson, S. F., & MacAlpine, G. M. 1989, AJ, 98, 1959
Francis, P. J., Hewett, P. C., Foltz, C. B., Chaffee, F. H., Weymann, R. J., & Morris, S. L. 1991, ApJ, 373, 465
Hawkins, M. R. S. 1986, MNRAS, 219, 417
Hewett, P. C., Foltz, C. B., Chaffee, F. H., Francis, P. J., Weymann, R. J., Morris, S. L., Anderson, S. F., & MacAlpine, G. M. 1991, AJ, 101, 1121
Hook, I. M., McMahon, R. G., Boyle, B. J., & Irwin, M. J. 1991, in PASP Conf. Series 21, The Space Distribution of Quasars, ed. D. Crampton, 67
Kendall, M. 1975, Multivariate Analysis (London: Griffin)
Kinney, A. L., Rivolo, A. R., & Koratkar, A. P. 1990, ApJ, 357, 338
Malkan, M. A., & Sargent, W. L. W. 1982, ApJ, 254, 22
McDowell, J. C., Elvis, M., Wilkes, B. J., Willner, S. P., Oey, M. S., Polomski, E., Bechtold, J., & Green, R. F. 1989, ApJ, 345, 13
Mittaz, J. P. D., Penston, M. V., & Snijders, M. A. J. 1990, MNRAS, 242, 370
Morris, S. L., Weymann, R. J., Anderson, S. F., Hewett, P. C., Foltz, C. B., Chaffee, F. H., Francis, P. J., & MacAlpine, G. M. 1991, AJ, 102, 1627
Murtagh, F., & Heck, A. 1987, Multivariate Data Analysis (Dordrecht: Reidel)
Robertson, J. G. 1986, PASP, 98, 1220
Smetanka, J. J., Bershady, M. A., Kron, R. G., Munn, J. A., Koo, D. C., & Majewski, S. R. 1991, in PASP Conf. Series 21, The Space Distribution of Quasars, ed. D. Crampton, 100
Sulentic, J. W. 1989, ApJ, 343, 54
Weymann, R. J., Morris, S. L., Foltz, C. B., & Hewett, P. C. 1991, ApJ, 373, 23
Whitney, C. A. 1983s, A&AS, 51, 443
———. 1983b, A&AS, 51, 463
Wilkes, B. J. 1986, MNRAS, 218, 331
Wills, B. J., Fang, D., & Brotherton, M. S. 1992, in Proc. 1991 Heidelberg Conference, Physics of Active Galactic Nuclei, ed. S. Wagner & W. Duschl, in press
Zamorani, G., Marano, B., Mignoli, M., Zitelli, V., & Boyle, B. J. 1992, MNRAS, in press