

Microsoft Excel ソルバーによる線形計画法の 滴定曲線解析への応用

吉村 季織^{a*}, 岡崎 正規^a, 中川 直哉^b

^a 東京農工大学大学院生物システム応用科学研究所, 〒184-8588 小金井市中町 2-24-16

^b 電気通信大学, 〒182-8585 調布市調布ヶ丘 1-5-1

*e-mail: yosimura@cc.tuat.ac.jp

(Received: January 6, 2003; Accepted for publication: February 17, 2003; Published on Web: April 22, 2003)

Microsoft Excel に添付されている最適解の検索ツール、ソルバーを用いて滴定曲線から試料溶液の pK 分布を見積もった。pK 分布は pK と濃度の関係を示し、不均一表面などの分析に用いられている。滴定曲線は、クエン酸水溶液を水酸化ナトリウム溶液で滴定することによって得た。滴定曲線の式を線形で表し、Excel ソルバーの線形計画法によって実験値にフィッティングした。Excel ソルバーによって見積もられた pK 分布は、シンプレックス法を用いた自作プログラムで見積もった pK 分布と比べても遜色がないものであった。pK 分布の見積もりに、Excel ソルバーを十分に活用できることを示すとともに、すべての Excel ユーザが気軽に pK 分布を見積もることができることを示している。

キーワード: Excel ソルバー, 線形計画法, 滴定曲線, pK 分布

1 緒言

パーソナルコンピュータ上で用いられている表計算ソフトウェアの中でも、Microsoft Excel (以下 Excel) は現在最も広く利用されている。Excel には通常の表計算ソフトとしての機能以外にも、統計ツールやデータベース機能など様々な機能が付加されている。ソルバーもそのひとつであり、専門的な知識を持たなくても最適化を行うことが可能となる。

カーブフィッティングを行うとき、しばしば最小二乗法 (LS) が用いられている。前報では、ソルバーによる非線形最小二乗法による滴定曲線の解析、pK 及び濃度の算出を行った [1]。LS では、誤差の二乗和を最小化させるが、誤差の絶対値の和を最小化させることで、最もデータに近づいた予測曲線を見出すことができる。予測曲線の理論式が線形で表されるなら、このフィッティングの問題は線形計画法 (LP) を用いて解くことができる。

本稿では、まず LP によるカーブフィッティングを、Excel ソルバーを用いて解く方法について解説する。

さらに応用として、最近の研究でも用いられている、pK 分布 [2–5] を作成する。

2 LP によるカーブフィッティング

LP とは、目的関数、制約式がともに線形で表される場合である。一般的な場合を示すと、変数 v_i に関して、

$$\text{目的関数: } f = \sum_{i=1}^m b_i v_i \text{ の最大化}$$

$$\text{制約式: } \sum_{i=1}^m (a_{ji} v_i) \leq c_j, \quad (j = 1, 2, \dots, n) \quad (1)$$

となる。 a_{ji} 、 b_i 、 c_j は係数。最大化問題は、目的関数の右辺の符号と、制約式の不等号を逆転させることで最小化問題となる。LP で解くときにまず考えなくてはならないのは、いかにして解きたい問題を線形で表すかである。

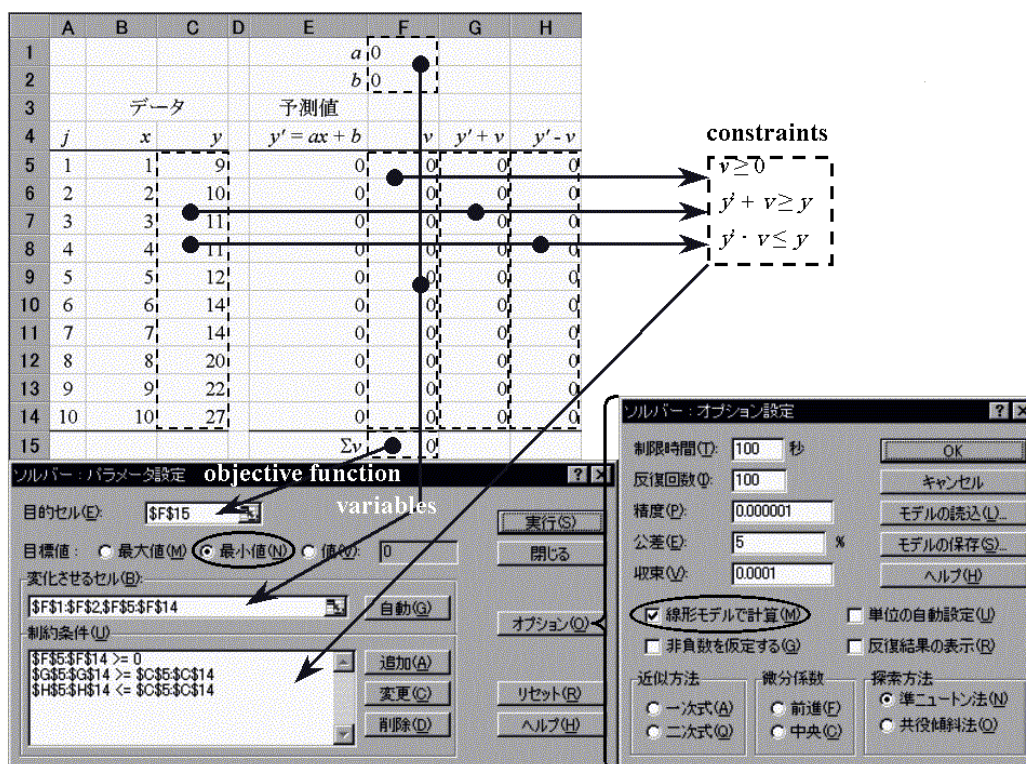


Figure 1. Preparation to solve a linear regression problem by LP in Excel solver.

m 個の独立変数 (x_1, x_2, \dots, x_m) からなる関数 $y' = \sum(c_i x_i)$ の場合を考える。 c_i は係数である。 j 番目の実際のデータ $(x_{j1}, x_{j2}, \dots, x_{jm}, y_j)$ と関数 y' による予測値 y'_j の誤差の絶対値の和 $\sum|y'_j - y_j|$ を最小化すれば、カーブフィッティングができる。しかしながら、この和は線形ではない。ここで、誤差の大きさを v_j とし、変数として扱うことにより、上記問題は以下のように書き換えられる。

$$\text{目的関数: } f = \sum_{j=1}^m v_j \text{ の最小化}$$

$$\begin{aligned} \text{制約式: } & y'_j + v_j \geq y_j \\ & y'_j - v_j \leq y_j \\ & v_j \geq 0 \quad (j = 1, 2, \dots, n) \end{aligned} \quad (2)$$

ここで、 v_j は実測値と予測値の誤差を直接求めているのではなく、あくまで変数であり、制約式によって誤差の大きさになるよう調整されている。目的関数 f は、明らかに線形である。関数 y' は線形なので、すべての制約式もまた線形である。ゆえにこの問題は LP によって解くことができる。この場合、関数 y' に含まれるすべての c_i と v_j が変数となる。目的関数の中に

は c_i が無いが、これは c_i の係数がすべて 0 であると考えるとよい。この問題を、たとえば C 言語の関数で解くには、その関数で解くことのできる入力形式や、式 (1) のように一般化された形にする必要がある。特に行列の生成は面倒である。しかしながら、Excel ソルバーを用いれば行列のことを考えることなく、ほとんど式 (2) の状態のまま問題解決まで進めることができる。

最もよく用いられる線形回帰への LP の適用を例としてあげる。まず、Figure 1 の左上に示すようなワークシートを用意する。セル B5 から C14 はこれからフィッティングするためのデータである。F1 および F2 には、回帰直線の係数 a, b とし、初期値として 0 を与えた。E5 から E14 は y の予測値 y' で、たとえば E5 には “= \$F\$1 * \$B\$5 + \$F\$2 ” が入力されている。F5 から F14 はデータとの誤差 v で、式ではなく実際に値を入れなくてはならないが、LP では変数として扱われるので、初期値として 0 に設定した。また、F5 から F14 間での合計値 (=SUM(F5:F14)) を F15 に入力した。これが、目的関数に相当する。G 列、H 列はそれぞれ制約式に当たる。5 行目を例に挙げると、セル内の記述内容は、それぞれ “=E5+F5”、“=E5-F5” である。

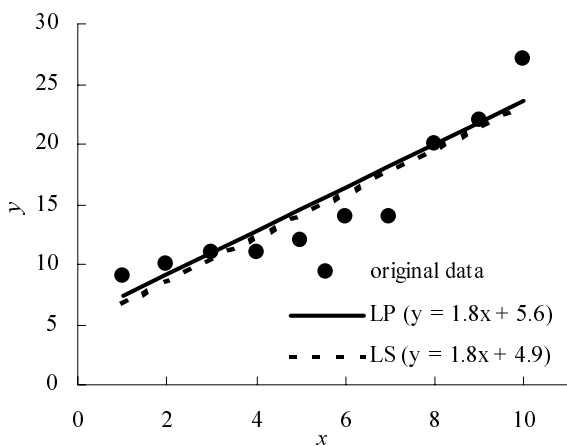


Figure 2. Comparison of results from linear regression by LP

次に、回帰のためのソルバーの設定を行う (Figure 1 下段)。ソルバーの設定とワークシートの関係を、矢印で示した。最小化の目的となるのは、 Σv だから、目的セルには“\$F\$15”を入力する。最小化問題であるから、目標値は“最小値”とした。変数は、回帰式の係数とすべての v なので “\$F\$1:\$F\$2,\$F\$5:\$F\$14”を入力した。10 個のデータがあるので、30 個の制約条件が存在するが、同一内容の条件については Figure 1 に示したように、“\$G\$5:\$G\$14 \geq \$C\$5:\$C\$14”のようにまとめて設定することができる。ここで、実行することで最適化は可能であるが、ソルバーには線形計画法を解くためのモジュールが搭載されているので、今回の場合それを利用した。パラメータ設定で、“オプション”を押すと、Figure 1 右図のようなオプション設定がでてくる。この中の、“線形モデルで計算”をチェックした。その他の設定は変更しなかった。これで設定は終了で、後は実行すればよい。

実行した結果、回帰係数 a 、 b それぞれ、1.8、5.6 が得られた。Figure 2 にデータと得られた回帰直線 (実線) の関係を示した。比較として最小二乗法による回帰結果も示した。これらの 2 本の直線は、わずかに異なっているが、よく似た結果となっている。データのばらつきが大きくなるにつれて、2 本の差も大きくなっていく。しかし、回帰の方法が異なっているので、どちらがより勝っている方法かを示すものではない。

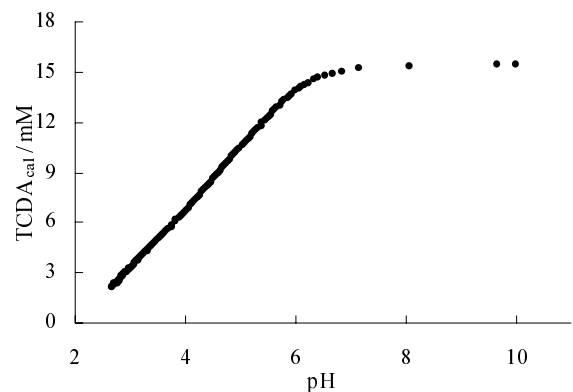
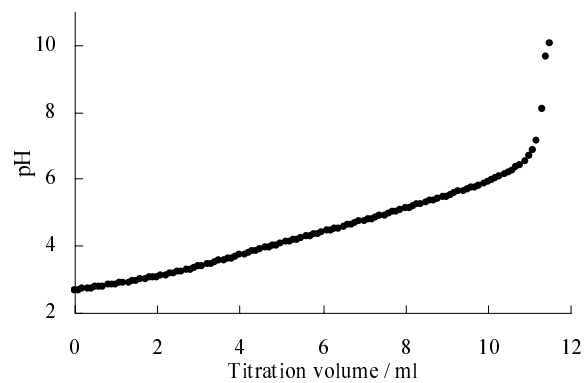


Figure 3. Original titration curve(upper) and converted titration curve(lower) for 5.12 mM citric acid.

3 LPによる滴定曲線解析

3.1 滴定曲線の線形変換

前述したように LP で問題を解くためにまず必要なことは、問題を線形で表すことである。一般に用いられる滴定曲線は (滴定体積) - (pH) で表されるが、これは非線形である。そこで、LP で扱えるようにするために、解離酸総濃度 (total concentration of dissociated acids : TCDA) [6] を導入し、以下の式にしたがって、滴定曲線を (pH) - (TCDA) 形式に変換する。

$$TCDA_{exp} = \frac{([H] - [OH]) (V_a + V_b) + c_b V_b}{V_a} \quad (3)$$

ここで、 V_a は試料溶液の初期体積、 c_b 、 V_b はそれぞれ滴定剤の濃度と、滴下体積である。(3) 式で求められる TCDA は実験で得られる値のみを用いるので、 $TCDA_{exp}$ と表す。

	A	B	C	D	E	F	G	BX	BY	BZ	CA	CB
1		c_b	0.135 M		i	1	2	71				
2		V_a	100 ml		pK_i	0	0.2	14		Σv		
3					c_i	0.00E+00	0.00E+00	0.00E+00		0.00E+00	TCDA _{cal}	TCDA _{cal}
4	V_b	pH	[H]-[OH]	TDA _{exp}		c_1X_1	c_2X_2	$c_{71}X_{71}$	TCDA _{cal}	v	$+v$	$-v$
5	0	2.68	2.09E-03	2.09E-03		0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00
6	0.1	2.69	2.04E-03	2.18E-03		0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00
120	11.5	10.03	-1.07E-04	1.58E-02	copy	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00

$= (C5 * (\$C\$2 + A5) + \$C\$1 * A5) / \$C\2
 $= F\$3 / (1 + 10^{(F\$2 - \$B5)})$
 $= \text{SUM}(F5:BX5)$
 $= \text{SUM}(BZ5:BZ120)$

Figure 4. Preparation of work sheet to determine the concentrations and pK values of citric acid.

次に、多価酸は異なる多数の一価酸の等濃度混合物として見なされ、それぞれの一価酸の解離定数は多価酸の逐次解離定数と等しくなる、という仮定から TCDA を理論的に導く (TCDA_{cal}) と、

$$\begin{aligned}
 \text{TCDA}_{cal} &= \sum_{i=1}^N c_i X_i \\
 X_i &= \frac{c_i}{1 + 10^{pK_i - \text{pH}}} \quad (4)
 \end{aligned}$$

となる。ここで、 c_i は pK_i を持つ酸の濃度である。式 (4) は pH に関しては非線形であるが、 pK_i をあらかじめ与えることによって、 X_i に関して線形と見なすことができる。すなわち、あらかじめ適当な pK_i を与えて、TCDA_{exp} を y 、TCDA_{cal} を y' に置き換えてやれば、式 (3) と (4) は式 (2) で表される LP にすることができる。この LP を解くことで最適な c_i の組を見つけることができる。

しかしながら、未知試料中に含まれている酸の種類数と、それぞれの酸の pK が一般に不明である。このようなときは、適当な範囲の pK を適当な間隔で区切ってあたえてやるとよい [6]。一般に (強酸、強塩基以外の) 酸の pK は、0 から 14 の間に存在する。また、一般に得られる滴定曲線もこの pK の範囲を外れた酸の正確な情報を含んでいない。そのため、 pK の範囲を 0 から 14 とした。また間隔を 0.2 とした。すなわち $pK_1 = 0.0, pK_2 = 0.2, \dots, pK_{71} = 14.0$ である。

3.2 滴定実験

5.12 mM のクエン酸水溶液 100ml を 135 mM の水酸化ナトリウム水溶液で滴定した。滴定体積は 1 滴あたり 0.10ml とし、pH10.03 まで行い 116 組の実験値を得た。

Figure 3 に滴定曲線と (pH)-(TCDA) 型に変換した滴定曲線を示した。

3.3 LP のための準備

Figure 4 に滴定曲線を解析するためのワークシートの一部を示す。A 列及び B 列の 5 行目以降が実験値で、その値をもとに D 列で TCDA_{exp} を算出した。2 行目 F 列から BX 列までは設定したそれぞれの pK を入力し、それぞれの列に対応する 3 行目がその酸の濃度とした。濃度は、LP において変数となるので、初期値を 0 とした。それぞれの列に対応する 5 行目以降で、実験によって得られた各 pH における $c_i X_i$ を算出した。絶対参照の使い方に注意すると、簡単に全体へコピーできる。それぞれの pH に関して、 $\Sigma c_i X_i$ すなわち TCDA_{cal} を BY 列 5 行目以降で計算した。BZ 列に TCDA_{exp} と TCDA_{cal} の差の絶対値 v とし、CA 列、CB 列でそれぞれ TCDA_{cal} + v 、TCDA_{cal} - v を計算した。最後に、BZ 列 3 行目で v の総和を計算した。

ソルバーの“パラメータ設定”及び“オプション設定”は、前述の例に少し変更を加えた。前述の例では、変数は回帰係数 a, b と誤差の絶対値 v_i であり、 v_i のみに非負制約を設定した。滴定曲線解析の場合、変数は回帰係数に当たる濃度 c_i と誤差の絶対値 v_i である。濃度は当然非負数のみをとるので、今回の例ではすべての変数に非負制約を設定する必要がある。そこで“パラメータ設定”の“制約条件”で設定せずに、“オプション設定”の“非負数を仮定する”をオンにした。これによってすべての変数が非負として扱われる。ここで注意しなければならないことは、Excel のソルバーで扱うことのできる変数の数が 200 までに制限されていることである。滴定曲線の解析では、滴定データの数と設定した pK の数の和が変数の数となる。本例で

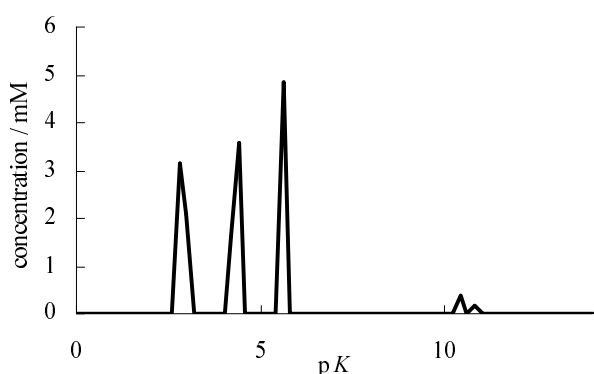


Figure 5. pK distribution of 5.12 mM citric acid solution

は、滴定データ 116 点、71pK なので、合計 187 変数でありソルバーで扱うことができる。もし 200 を越えるようならば、変数を減らす工夫をする必要がある。さらに“ オプション設定 ”において反復回数を 1000 に増やした。これは、初期値の 100 では最適化が終了せずに計算途中で継続するか否か確認させられるので、これを回避するために、計算精度には影響しない。

計算は、Windows NT 4.0 上で動作している、Excel 2000 にて行った。

4 結果と考察

計算により得られた数値結果を折れ線グラフとして、Figure 5 に示した。このような図を pK 分布、pK スペクトルなどと呼ぶ。Figure 5 より、pK3.0、4.5、5.5 付近にそれぞれ大きなピークが現れ、pK7 付近と、10 と 11 の間に小さなピークが見られた。前者の 3 本のピークがクエン酸由来で、それぞれクエン酸の逐次解離に対応する。後者の小さなピークは、炭酸などの不純物により現れたものと思われる。同じ滴定曲線をシンプレックス法を用いて作成した自作プログラム [7] で解析し、ソルバーと比較した。シンプレックス法は非負制約付線形計画法を解く最も一般的な方法である。前者 3 本のクエン酸由来のピークでは、ソルバー算出値と自作プログラム算出値間での相対誤差の平均が、pK では $3.10 \times 10^{-8} \%$ 、濃度では $8.71 \times 10^{-7} \%$ と非常に小さく有意なものではなかった。pK7 付近ピークでも 0.02 % 以下の誤差であり、有意とは言えない。pK10-11 のピークは、自作プログラムでは 1 本しか現れなかった。これは計算法によって pK が変化する不安定なピークであることから、比較ができなかった。この

Table 1. pK value and concentration of citric acid.

reference[12]	pK		concentration mM
	reference[12]	experimental	
2.90	2.88		5.25
4.34	4.34		5.18
5.66	5.60		4.87

ように、Excel ソルバーで解いた結果は、専用のプログラムで解いた結果と遜色がなかった。

それぞれのピークは、1 つ以上の pK とそれに対応する濃度 c によって構成されている。例えば、pK3.0 付近のピークは、pK2.8 の酸 3.16 mM と pK3.0 の酸 2.09 mM からなる。pK5.5 付近のピークが大きく見えるのは、このピークに含まれる pK が 1 つだけであるためである。また、ピークとして認められる pK 以外の領域では、すべての pK_i に対応する濃度 c_i は 0.0 mM であった。TCDA_{exp} と TCDA_{cal} の差の平均は、 1.37×10^{-2} mM と非常に小さく、良くフィッティングが行われていることが分かる。

pK の間隔が 0.2 と広いと、それぞれのピークも幅が広がった。しかしながら、1 種類の解離に対応する pK は理想的には 1 つであり、そのことを pK 分布で表すと、完全な 1 本線になるはずである。ピークの出た領域内をさらに細かい pK 間隔で区切り、同様の計算を繰り返すことで、ピークはより細くなり、より正確な濃度や pK を見積もることができるようになっていくことが明らかになっている [6]。このことは、ピークの現れている領域内に真の pK があることを示唆している。今回得られた結果は荒いものであるが、0.0 でない c_i に対応する pK_i を加重平均により補正することでより正確な pK を見積もることができる。

$$pK = \frac{\sum c_i pK_i}{\sum c_i} \quad (\text{ピークを構成する pK に関して}) \quad (5)$$

濃度は、ピークを構成する pK に対応する c の合計とする。

$$c = \sum c_i \quad (6)$$

これらの補正法によって得られた、クエン酸の pK と濃度を Table 1 に示した。補正によって算出された pK の値は、文献値に良く一致していることが分かる。調製した濃度は 5.12 mM であったから、やや誤差が大ききようである。この誤差は試料についての情報が全くなかったことや多価酸を 1 価酸の混合で近似したことよりも、 pH メーターの持つ誤差など、実験時の誤差の方が大きく影響している。しかしながら、この 3 つの酸の平均濃度は 5.10 mM と調製した濃度に充分近く、等量点の見積もりは正確に行うことができる。

5 終わりに

Excel ソルバーは、Excel ユーザーなら誰でも使うことのできる最適化ツールである。本稿では、ソルバーに内包されている線形計画法のモジュールを活用して、水溶液未知試料の滴定曲線より、 pK 分布を見積もる方法を示した。もしこの手順を自作プログラムとして作成するならば、線形計画法の解を求めるモジュールを作らなくてはならないほか、線形計画法に則るために目的関数の係数ベクトル、制約式の係数行列などを生成するためのルーチンを記述しなくてはならない。もちろんソルバーで問題を解くとすると、計算時間を犠牲にしなくてはならないが、滴定に要する時間に比べると十分に小さく、実験の中での律速段階にはならない。

本法は、最低限の試料の情報のみで、 pK 分布を見積もるので、特定の酸の定量や pK 決定など、精密な分析を行うことには向いていない。逆に、 pK 分布の傾向を見積もることができるため、未知試料や複雑な混合系の分析に適している。実際に、 pK 分布は酸化物 [8] や粘土鉱物表面 [9, 10] やイオン交換樹脂 [11]、微生物表面 [4, 5]、有機物 [3] など、不均一表面、複雑な系の分析に使われている。

最近のパソコンには、あらかじめ Excel がインストールされていることが多いため、本法はあらゆる

ユーザーに、 pK 分布を見積もる機会を提供するものである。

参考文献

- [1] 吉村季織, 岡崎正規, 中川直哉, *J. Chem. Software*, **7**, 191 (2001).
- [2] P. Brassard, J. R. Kramer and P. V. Collins, *Environ. Sci. Technol.*, **24**, 195 (1990).
- [3] D. S. Smith and J. R. Kramer, *Environ. Int.*, **25**, 307 (1999).
- [4] J. S. Cox, D. S. Smith, L. A. Warren and F. G. Ferris, *Environ. Sci. Technol.*, **33**, 4514 (1999).
- [5] I. Sokolov, D. S. Smith, G. S. Henderson, Y. A. Gorby and F. G. Ferris, *Environ. Sci. Technol.*, **35**, 341 (2001).
- [6] N. Yoshimura, M. Okazaki and N. Nakagawa, *Anal. Sci.*, **16**, 1331 (2000).
- [7] 吉村季織 未発表
- [8] C. Contescu, J. Jagiello, and J. A. Schwarz, *Langmuir*, **9**, 1754 (1993).
- [9] T. J. Bandosz, J. Jagiello and J. A. Schwarz, *J. Colloid Interface Sci.*, **182**, 570 (1996).
- [10] T. J. Bandosz, J. Jagiello and J. A. Schwarz, *J. Phys. Chem.*, **100**, 15569 (1996).
- [11] S. Y. Bratskaya and A. P. Golikov, *J. Anal. Chem.*, **53**, 234 (1998).
- [12] 日本化学会編, 化学便覧 基礎編 II, 改訂 4 版, 丸善 (1993), p.318.

Application of Linear Programming Method by Microsoft Excel Solver to Titration Curve Analysis

Norio YOSHIMURA^{a*}, Masanori OKAZAKI^a and Naoya NAKAGAWA^b

^aGraduate School of Bio-Applications and Systems Engineering, Tokyo University of Agriculture and Technology
2-24-16 Nakacho Koganei, Tokyo 184-8588, Japan

^bThe University of Electro-Communications, 1-5-1 Chofugaoka, Chofu, Tokyo 182-8585, Japan

**e-mail: yosimura@cc.tuat.ac.jp*

Using Microsoft Excel solver, which is an add-in program to solve optimization problems, a pK distribution was estimated. The pK distribution shows the relation between the pK values and concentrations of acids in a sample solution and is applied to analyze a heterogeneous surface. A titration curve was obtained from titration of a citric acid solution with a sodium hydroxide solution. The theoretical equation of the titration curve was converted into a linear equation. The equation was fitted to experimental data by the linear programming method included in Excel solver. The pK distributions estimated by Excel solver and by a specially designed program applying the simplex method are almost equal. This indicates that Excel solver can not only be applied for excellent estimation of pK distribution, but also that all Excel users can easily carry out similar determinations.

Keywords: Excel solver, Linear programming method, Titration curve, pK distribution

