

The Nonparametric Adaptive Grid Algorithm for Population Pharmacokinetic Modeling

W.M. Yamada^{a,f}, J. Bartroff^b, D. Bayard^f, J. Burke^c, M. van Guilder^f, R.W. Jelliffe^f, R. Leary^{d,1}, M. Neely^{e,f}, A. Kryshchenko^b, A. Schumitzky^{b,f,*}

^a*Department of Psychology, Azusa Pacific University, 901 E. Alosta Ave., Azusa, CA 91702-7000, USA*

^b*Department of Mathematics, University of Southern California, Los Angeles, CA 90089-2532, USA*

^c*Department of Mathematics, University of Washington, Seattle, WA 98195-4350, USA*

^d*San Diego Supercomputer Center, 9500 Gilman Drive, MC 0505, La Jolla, CA 92093-0505, USA*

^e*Pediatric Infectious Diseases, Children's Hospital of Los Angeles, Keck School of Medicine, University of Southern California, Los Angeles, CA 90027, USA*

^f*Laboratory of Applied Pharmacokinetics, Children's Hospital-Los Angeles, 4650 Sunset Blvd., Los Angeles, CA 90027, USA*

Abstract

The nonparametric adaptive grid (NPAG) algorithm is described. NPAG is a primal dual interior point method for determining the nonparametric maximum likelihood distribution of random parameters for a nonlinear regression model. In pharmacokinetic (PK) population modeling, NPAG is useful for the case where a series of measurements are made on diverse individuals, distinguished by diverse measurement conditions. This paper includes sufficient mathematical background and program detail to allow computational scientists to translate NPAG from PK

*Corresponding author. Tel.: +1 818-249-9444; Fax: +1 213-740-2424.

Email addresses: wyamada@apu.edu (W.M. Yamada), bartroff@usc.edu (J. Bartroff), dbay007@earthlink.net (D. Bayard), jvburke@uw.edu (J. Burke), uphill@cox.net (M. van Guilder), jelliffe@usc.edu (R.W. Jelliffe), Bob.Leary@certara.com (R. Leary), mneely@chla.usc.edu (M. Neely), chubatiu@usc.edu (A. Kryshchenko), schumitzky@aol.com (A. Schumitzky)

¹Present address: Pharsight, inc., Cary, NC 27518, USA

population modeling to other domain areas. Examples are given.

Keywords: Adaptive Grid; Nonparametric maximum likelihood; Primal dual interior point method; Population pharmacokinetics

1. Introduction

The question of how to describe patient variability in drug behavior is an ever present one, see e.g. [1] or [2]. In this paper we describe the estimation of a distribution of PK parameters given a structural PK model and a set of observations on a population of patients, where no assumption is made regarding the true distributions of the PK model parameters.

1.1. A brief history of NPAG

The development of the nonparametric maximum likelihood (NPML) method for estimating an unknown probability distribution F began in the 1980s. Lindsay [3] and Mallet [4] proved the fundamental result about the NPML estimator F^{ML} , namely that F^{ML} could be found in the set of discrete distributions with no more support points than the number of subjects N in the population. Both Lindsay's and Mallet's results were based on convexity theory. The problem remained to find the position of the support points of F^{ML} and the corresponding probabilities.

Schumitzky [5] derived a nonparametric EM algorithm for the F^{ML} called NPEM which was based on statistical theory. The NPEM algorithm was very simple: A large grid G^0 of potential support points was laid out on the domain of F . The NPML estimator of F^{ML} was then reduced to finding the corresponding probabilities of the support points and the corresponding likelihood $L(F^{ML}|G^0)$. This was done by the EM algorithm. Most of the support points had very small probabilities, less than 10^{-12} and were deleted from G^0 leaving a new and smaller

Table 1: Notation introduced in section 1; mostly describing patient data

$y_i \in Y$	$Y = \{y_1, \dots, y_N\}$ $y_i = \{y_{im}\}$ y_{im} is the i th patient's observation at time m
x_i	x_i is the i th patient's observation conditions
$v, f(x, \beta)$	f is the structural PK model that defines the expected response given x and β v is the Normally distributed noise term
$\Omega = [a_1, b_1] \times [a_2, b_2] \times \dots \times [a_k, b_k]$	Ω is an Euclidean box, defined by user-input minimum and maximum values for each parameter.
$\beta_i \in \Omega$	β_i is the i th patient's pharmacokinetic parameters.
F F^{ML}	F is a distribution on Ω F^{ML} is the Maximum Likelihood distribution, maximizing $p(Y F)$ over the space of all F
$G = \{(g_j, p(g_j)), j = 1, \dots, L\}$ $g_\Omega := \{g_j\} \in G_\Omega$ $g_A := \{g_j\} \in G_A$	G a discrete distribution on Ω expanded grid condensed grid
N	number of observed patients
L	number of points in G

grid G^1 . The process was repeated giving rise to a sequence of smaller grids G^k and larger likelihoods $L(F^{ML}|G^k)$. When the difference of likelihoods was sufficiently small the process was deemed to converge. By theory, the number of support points was not greater than the number of subjects in the population, and usually a lot less. In order to achieve good resolution, a large number of grid points G^0 must be chosen, which can lead to high computational demands requiring a large-scale parallel supercomputer. Further, the NPEM algorithm had linear convergence and was very slow; and the closer it got to convergence the slower it got.

In the late 1990s, two major improvements were made to the NPEM algorithm. Robert Leary (Pharsight Corporation) developed the Adaptive Grid method. In this approach, the original NPEM algorithm is applied to a modestly sized grid H^0 to produce an F^{ML} solution. Support points with very small probability are deleted from H^0 . Then new support points are now created by adding support points around each of these remaining support points in H^0 leading to a new grid H^1 and an improved likelihood. The process is repeated until convergence of likelihood, while gradually reducing the size of the refinement regions around each successive set of support points. The modified NPEM algorithm uses far fewer total grid points than the original NPEM to achieve the same accuracy of solution [6].

Around the same time, James Burke (University of Washington) developed a Primal Dual Interior Point algorithm for solving the NPML estimation problem which has near quadratic convergence, see [7]. Restricted to a fixed grid it could replace the EM algorithm of the Leary Adaptive Grid NPEM program. Leary and Burke put the two programs together resulting in what is now essentially the

NPAG program. NPAG is 100-1000 times faster than NPEM.

NPAG is written in Fortran code but the primal dual interior point method can be found in two elegant Matlab programs written by James Burke and Bradley Bell [8, 9]. The current user interface to NPAG is Pmetrics [10], which is a package written in R and downloadable from www.lapk.org. The primary purpose of this paper is to present a thorough description of the NPAG algorithm.

1.2. Statement of the Problem

Traditionally, PK parameter estimation begins by assuming that each patient observation is described by

$$y_i = f(x_i, \beta_i) + v_i, \quad i = 1, \dots, N \quad (1)$$

the $\{y_i\}$ are independent random vectors describing the observed measurements for the i th patient; the function f is a structural PK model, which describes the expected drug concentration profile given x_i and β_i ; x_i is a vector of values that include the initial concentrations, dose schedule, measurement schedule, and the covariate values for the i th patient; β_i is a random vector of pharmacokinetic parameters specific for the i th modeled subject, v_i is an assumed normal error model with mean 0 and covariance Σ_i .

The values of the $\{\beta_i\}$ are assumed to belong to a closed and bounded set Ω in Euclidean space. Further the $\{\beta_i\}$ are assumed to be independent and identically distributed random vectors with common but unknown distribution F on Ω . The PK population problem is to estimate F given the data $Y = \{y_1, \dots, y_N\}$. It is assumed that $f(x_i, \beta_i)$ is a continuous function of β_i on Ω for all $i = 1, \dots, N$.

From equation (1), the Likelihood Function is given by:

$$p(Y|F) = \prod_{i=1}^N \int p(y_i|\beta_i) dF(\beta_i) \quad (2)$$

The Maximum Likelihood distribution F^{ML} maximizes $p(Y|F)$ over the space of all distributions defined on Ω .

It is important to note that the optimization of equation 2 with respect to the distribution F can also be viewed as an Empirical Bayes method to estimate the conditional probability of $(\beta_1, \dots, \beta_N)$ given the data (Y_1, \dots, Y_N) . This idea has gained considerable attention recently, see [11].

The NPAG algorithm is motivated by the theorems of [3] and [4] which state, under reasonable hypotheses, that F^{ML} can be found in the class of discrete distributions with at most N support points, which we write as:

$$F^{ML}(\beta) = p(g_1)\delta_{g_1}(\beta) + \dots + p(g_K)\delta_{g_K}(\beta), \quad K \leq N$$

where $\{g_j\}$ are the support points and $\{p(g_j)\}$ are the non-negative weights which sum to one; δ_g represents the delta distribution on Ω with the defining property $\int h(\beta)d\delta_g(\beta)d\beta = h(g)$ for any continuous function $h(\beta)$ on Ω ; or more generally $\int h(\beta)dF^{ML}(\beta) = \sum_{j=1}^K p(g_j)h(g_j)$. This property of transforming integration into summation is crucial for multiple model type controllers, see [12]. Further it is proved in [13] that F^{ML} is a consistent estimator of the true distribution F , i.e. as $N \rightarrow \infty$, F^{ML} converges in distribution to F .

We begin discussion in the next section with an example that considers the effect on estimating a population distribution that contains PK parameters with unspecified distribution, and an assumed normally distributed measurement noise v , with mean 0 and standard deviation σ .

NPAG is a nonparametric maximum likelihood (NPML) method that estimates the patient population distribution using a discrete distribution of support points and their probabilities, $G_A = \{(g_j, p(g_j)), j = 1, \dots, L\}$. NPAG determines F^{ML} by searching in Ω for the $\{g_j\}$ corresponding to F^{ML} . Search is facilitated by

the interior point method (IPM) of [7], which is employed to solve the following maximization problem,

$$\max_{G_A} \log \prod_{i=1}^N \sum_{j=1}^L p(y_i|g_j)p(g_j) \quad (3)$$

on a fixed grid. Iteratively expanding the grid, and contracting it to the high probability points of equation (3) at each iteration converges to F^{ML} .

The NPAG algorithm differs from the earlier NLME approach of [14] in that no assumption of the underlying shape of the distribution of PK parameters is made at any stage of the optimization. The importance of using modeling approaches freed from the assumption of normally distributed PK parameters has been accepted for over a decade [15, section VII.C]. Lack of parametric constraint permits easier elucidation of unsuspected subpopulations, see e.g. [16], without imposing an assumed parametric form on each subpopulation, or potentially hiding correlations due to shrinkage induced by using FO or FOCE posthoc β values as a prior to a nonparametric optimization equation [17, 18]. However, from drug trial and patient care perspectives, the primary motivation for using NPML techniques is their ability to generate models with strongest predictive power given a set of observed data [19, 20]; and therefore, require the least degree of therapeutic drug monitoring (TDM) during drug dosage individualization. Also, individualization is easily accomplished via multiple model design using the discrete distributions generated by optimal nonparametric methods such as NPAG [21].

Optimization of equation (3) requires maximization over *both* placement and probability of support points in Ω . Placement of points is determined by repeating the procedure of adding new points to $\{g_j\}$, calculating their probabilities, and culling the low probability points. The procedure of expanding the current estimate of F^{ML} and condensing to a new estimate is described in section 4. The ex-

pansion/condensation search through Ω is guided by calculation of support point probabilities, which is carried out by the interior point method of [7], described in section 3. We conclude in section 5, with several remarks on the practicality of using NPAG over other algorithms for population PK modeling.

2. Motivating Example

The parametric model, expressed in equation (4) below, is the system used here to simulate 1000 fictitious patients, 500 used for modeling and 500 used for testing. The random variables, K_{el} and vol_c , are independent. vol_c and K_{el} are stated exactly in equation (4). K_{el} describes a population of both fast and slow eliminators of the drug. The two distributions of K_{el} both have the same variance, and the modes of the mixture are approximately $3 \times \sigma$ apart. Each simulated patient was given an intravenous (IV) drug infused of $600mg$ over $0.5hr$ every $24hrs$, for 10 days. Four measurements were made for each patient: the peak and trough after the first dose, and the peak and trough after the 10th dose. No process or assay noise was modeled during simulation: all variability in patient response to the drug was due to the randomness of the population PK parameters. The dosage regimen is notated $IV(t)$. All patients in the population had the same covariate measurements, so that $x = \{IV(t)\}$; and $\beta = (K_{el}, vol_c)$,

$$\begin{aligned}
 f(x, \beta) &= A/vol_c & (4) \\
 \text{where, } \frac{d}{dt}A &= IV(t) - K_{el}A \\
 K_{el} &\sim 0.5 \left(\mathcal{N}\left(0.0722, 5.213 \times 10^{-5}\right) \right. \\
 &\quad \left. + \mathcal{N}\left(0.101, 5.213 \times 10^{-5}\right) \right) \\
 vol_c &\sim \mathcal{N}(17.5, 19.141)
 \end{aligned}$$

$r \sim \mathcal{N}(m, \sigma^2)$ means r is normally distributed with mean m and variance σ^2 .

The means of the PK distributions describe a 70kg person (volume equal to 17.5L), who is given a drug with half life $T_{1/2} \approx 8hrs$ (K_{el} of 8.66434×10^{-2}). However, the population consists of both fast and slow eliminators with mean $T_{1/2} = 6.857hrs$ or $9.6hrs$. A two-dimensional scattergram of the PK parameters used to simulate each patient is graphed using grey diamonds in figure 1.

2.1. NPAG dependence on residual error

Health agency industry guidelines for PK analysis require discussion of the role of residual error [15, 22]. In this section we use NPAG to make a population model of the first 500 simulated patients under conditions of increasing assumed residual error. Actual simulation error was zero: administered amounts, measured amounts, and timing of drug doses and measurements, were all exact. For the optimizations here, we assumed a normally distributed error with $\mu = 0$ and $\sigma = 0.1, 0.2, 0.4, 0.8, 1.6,$ or 3.2 . We simulated four assay measurements for each patient: the peak and trough data for the first and the last doses. Because all patients began with no drug, and there was only one PK model compartment, these four points represent the most extreme fluctuations of the drug concentration profile.

Since there is no noise in the simulated model population, for the case of a near zero assumed noise, intuition suggests that a population modeling algorithm should ‘find’ approximately one point for every unique subject in the population. This follows simply because no noise implies that every distinct measurement must arise from a unique set of PK parameters. For the case where $\sigma = 0.1$ NPAG found 450 support points for the 500 simulated subjects (see figure 1), that correspond well to the MC generated subjects (pairs of parameter values).

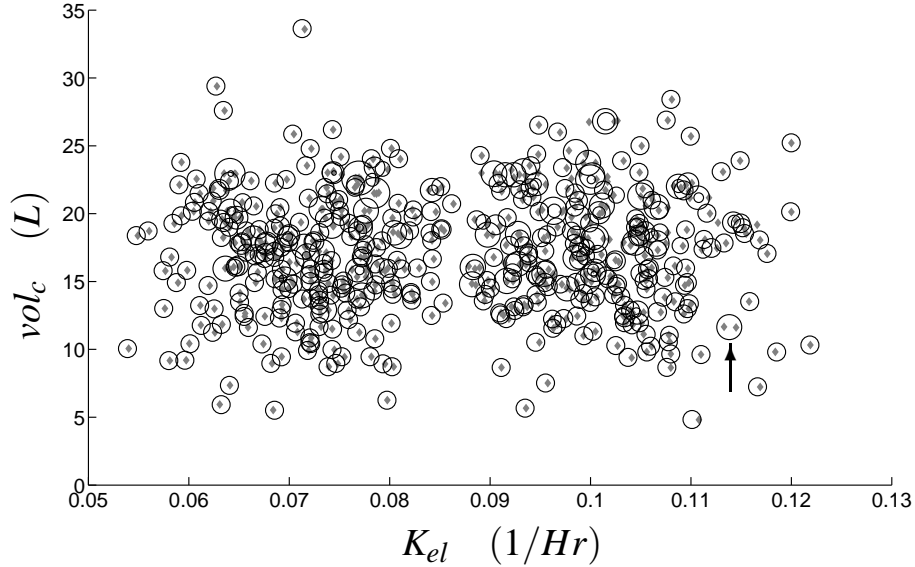


Figure 1: The scattergram of true individual parameters used during Monte Carlo simulation are plotted in grey diamonds. The scattergram of the NPAG population model assuming an assay error $v_i \sim \mathcal{N}(\mu = 0, \sigma = 0.1)$ are plotted in circles. There are 450 points in the NPAG model, with probabilities on the range (0.0046%, 0.8%). Each point is plotted with diameter relative to its probability. The upward arrow indicates one of the support points with probability 0.8%, explained in text.

To demonstrate the effect of increasing noise, we first rearrange equation (1) as follows.

$$v_i = y_i - f(x_i, \beta_i) \quad i = 1, \dots, N$$

and note that during optimization we assume v_i are normally distributed with a standard deviation determined only by the level of assumed measurement error. Increasing the suspected assay noise is equivalent then to increasing the uncer-

tainty of measurement. Increased uncertainty implies decreased ability to distinguish patient observations resulting from similar PK values. NPAG manifests this uncertainty by decreasing the number of distinct support points in the population description. This intuition is shown to have mathematical support by [3] and [4], who apply the theorem of [23] to show that the number of support points produced by the NPAG algorithm is always less than or equal to the number of modeled subjects.

Reduction of support points begins with a level of assumed noise as little as $\sigma = 0.1$. Careful inspection of figure 1 ($\sigma = 0.1$ condition) reveals several support points of greater likelihood that are placed where two or more individuals have similar parameter values. For example, an upward arrow points to a support point with probability 0.4% near $K_{el} = 0.1138$ and $vol_c = 11.643$, that encompasses two patients with PK values of (0.1145, 11.61) and (0.1133, 11.66), respectively. The reduction of support points as a function of assumed assay error during optimization is further demonstrated in figure 2.

2.2. Estimating mean population response

The NPAG distribution F^{ML} may be used for prospective study of a new patient. The expected concentration measurements for the $(i + 1)st$ patient are

$$E[f(x_{i+1}, \beta_{i+1})] = \int f(x_{i+1}, \beta) dF^{ML}(\beta) \quad (5)$$

500 new patients were simulated using Monte Carlo methods as described in section 2. Residual error statistics are given in table 2.

The use of NPAG to model a novel patient, especially as therapeutic drug monitoring data becomes available, is discussed in [24].

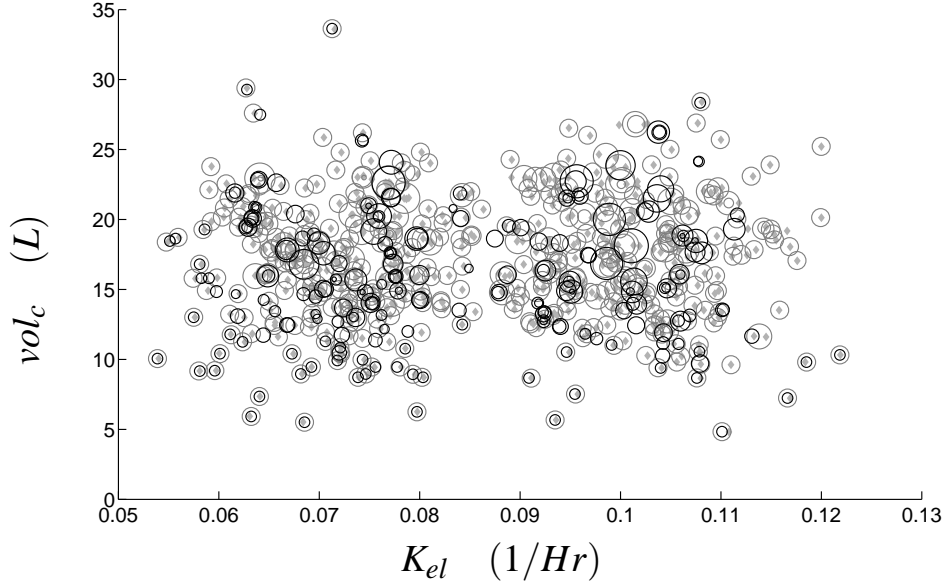


Figure 2: The scattergram of the NPAG distribution is plotted for $\sigma = 0.4$ (black circles) and for $\sigma = 0.1$ (grey circles). The higher assumed error results in the number of support points being further reduced to 217. The true individual parameters are plotted in grey diamonds.

3. Mathematical Description of NPAG

We now describe the primal dual interior point method as it applies to our NPAG program. Since this algorithm is written in Fortran computer code, we give a mathematical description in what follows. This material is standard in Convex Optimization Theory so no proofs will be given. An excellent reference is the book by [25].

From equation (2), the log likelihood is given by

$$\log p(Y|F) = \log \prod_{i=1}^N \int p(y_i|\beta_i) dF(\beta_i) = \sum_{i=1}^N \log \sum_{j=1}^L p(y_i|g_j) p(g_j)$$

Table 2: Statistics of Residual Errors. Using the NPAG model produced under varying assumptions of noise level, equation (5) was used to model a new population ($N = 500$). Mean and standard deviation of the residual errors ϵ are reported here. The population observation means and standard deviations are; First Peak: 17.90, 5.93; First Trough: 2.634, 1.423; 10^{th} Peak: 20.79, 7.185; 10^{th} Trough: 6.866, 3.212

	$\sigma = 0.1$		0.2		0.4	
	μ_{ϵ}	σ_{ϵ}	μ_{ϵ}	σ_{ϵ}	μ_{ϵ}	σ_{ϵ}
First Peak	0.0183	0.6229	0.0761	0.7449	0.1518	0.7363
First Trough	-0.0022	0.2032	0.0034	0.2238	0.0007	0.2638
10^{th} Peak	0.0293	0.6837	0.0940	0.7966	0.1675	0.7842
10^{th} Trough	0.0121	0.4123	0.0281	0.4548	0.0300	0.5293
	$\sigma = 0.8$		1.6		3.2	
	μ_{ϵ}	σ_{ϵ}	μ_{ϵ}	σ_{ϵ}	μ_{ϵ}	σ_{ϵ}
First Peak	0.1057	0.7808	0.1945	1.1023	0.5730	1.723
First Trough	-0.0036	0.3370	-0.0322	0.5454	0.0146	0.8501
10^{th} Peak	0.1236	0.8281	0.1970	1.1518	0.6568	1.8634
10^{th} Trough	0.0176	0.6658	-0.0289	1.0648	0.1035	1.6527

The Maximum Likelihood distribution F^{ML} maximizes $\log p(Y|F)$ over the space of all distributions defined on Ω . The theorems of [3] and [4] state that F^{ML} can be found in the class of discrete distributions with at most N support points.

Table 3: Notation used to describe the optimization problem.

$g_A \subset g_\Omega$	g_A are the high probability supports of G_Ω . g_A are those points corresponding to active constraints, i.e. $\xi \rightarrow 0$.
$\Psi_{i,j}$	$\Psi_{i,j} = p(y_i g_j)$ assuming normal error, equation (26).
\mathcal{P}	Primal problem, equation (7)
\mathcal{D}	Dual problem, equation (8)
\mathcal{P}_ρ	The program to solve \mathcal{P}
$\phi(z)$	Equation 6

Consequently to find F^{ML} it is sufficient to maximize equation (3) over the set of all $G_A = \{(g_j, p(g_j)), j = 1, \dots, L\}$.

Let $\Psi_{i,j} = p(y_i|g_j)$, and $\lambda_j \propto p(g_j)$, and assume λ_j have been normalized. Note that Ψ is an $N \times L$ dimensional matrix, and $\sum_{j=1}^L p(y_i|g_j)p(g_j) = (\Psi\lambda)_i$. We assume that $\Psi e_L > 0$, for e_L a L -length column vector with components all equal to 1, i.e. the row sums of Ψ are strictly positive. Now for any L -dimensional vector $z = (z_1, \dots, z_L)$, define the function

$$\phi(z) = \begin{cases} -\sum_{j=1}^L \log z_j & \text{if } z_j > 0 \\ +\infty & \end{cases} \quad (6)$$

Now fix the support points in G_A . The problem of maximizing equation (3) w.r.t. the weights $\{p(g_j)\}$ is equivalent to the following problem.

$$\underset{\lambda}{\text{minimize}} \quad \phi(\Psi\lambda) \quad \text{subject to } e^T \lambda = 1, \lambda \geq 0$$

Table 4: Notation used in the IPM.

λ	λ is the L -length primal variable of the interior point method λ is proportional to $p(g)$, equation (7)
z	z is the N -length primal variable of the second constraint of the KKT conditions, equation (9). Note that $z \rightarrow P$.
ξ, w	ξ and w are the L -length, and N -length IPM dual variable arrays
$\rho = \varphi\mu$	ρ is the homotopy relaxation parameter of the IPM. $\mu = \lambda^T \xi / L$ See equations (17) , 23, and 25
e_n	a n -length ones column array; subscript is omitted where length is clear from context
$\Lambda, \Xi, W, Z, \mathbf{P}$	diagonal matrices constructed from λ, ξ, w, z , or P

First for convenience we include the constraint $e^T \lambda = 1$ in the problem as a penalty to the objective function.

$$\text{Primal Problem, } \mathcal{P} \quad \underset{\lambda}{\text{minimize}} \quad \phi(\Psi\lambda) + N(e^T \lambda - 1) : \quad \lambda > 0 \quad (7)$$

the resulting problem has the same optimal solution as before [7, Proposition 4.2].

The Fenchel-Rockafellar convex dual program is

$$\text{Dual problem, } \mathcal{D} \quad \underset{w}{\text{minimize}} \quad \phi(w) \quad \text{subject to } w \geq 0, \Psi^T w \leq Ne \quad (8)$$

From [7, theorem 4.6] solutions to \mathcal{P} and \mathcal{D} exist, with solution to \mathcal{D} unique.

Furthermore, λ solves \mathcal{P} and w solves \mathcal{D} if and only if the Karush-Kuhn-Tucker (KKT) conditions hold

$$\begin{bmatrix} \Psi^T w + \xi - N e_L \\ W \Psi \lambda - e_N \\ \Lambda \Xi e_L \end{bmatrix} = 0 \quad (9)$$

where ξ is a non-negative L -dimensional vector of slack variables, and $\Lambda = \text{diag}(\lambda)$, $\Xi = \text{diag}(\xi)$, and $W = \text{diag}(w)$.

Equation 9 may be simplified further using the following change of variables

$$w = w/N$$

$$\xi = \xi/N$$

$$\lambda = N\lambda$$

This has the effect of changing the first KKT condition to

$$\Psi^T w + \xi = e_L$$

while leaving the other conditions unchanged.

3.1. Interior Point Method on a Fixed Grid: the central path

Algorithm 1 Interior Point Method (IPM). Initialization is described in section 4.4. Some numerical details are omitted here for clarity, these are found in [8]

Repeat steps 1 through 5

1. Solve for Δw . Specific equations are given in equations 20 and 19.

$$\mathbf{H}\Delta w = \mathbf{rhs} \rightarrow \Delta w \quad (10)$$

2. Update the dual and primal variables

$$\Delta \xi_j = -\Psi^T \Delta w \quad (11)$$

$$\Delta \lambda_j = \frac{\rho}{\xi_j} - \lambda_j - \Delta \xi_j \frac{\lambda_j}{\xi_j} \quad (12)$$

3. Calculate the damping factors of the Newton step

$$\alpha' = \min \left(1^-, \frac{-1}{\min \left(\frac{-1}{2}, \frac{\Delta \lambda_j}{\lambda_j} \right)} \right), \text{ and } \alpha'' = \min \left(1^-, \frac{-1}{\min \left(\frac{-1}{2}, \frac{\Delta \xi_j}{\xi_j} \right)} \right) \quad (13)$$

4. Calculate the new values of λ , ξ and w

$$\lambda : \lambda_j^+ \rightarrow \lambda_j + \alpha' \Delta \lambda_j, \quad (14)$$

$$\xi : \xi_j^+ \rightarrow \xi_j + \alpha'' \Delta \xi_j,$$

$$w : w_i^+ \rightarrow w_i + \alpha'' \Delta w_i.$$

Interior Point Method, continued.

5. Calculate the boundary and test conditions.

$$\mu = \frac{\lambda^T \xi}{L}$$

$$r = |e_N - W\Psi\lambda| \quad (15)$$

$$\mathbf{gap} = \left| \frac{\sum_i \log(w_i) + \sum_i \log(\Psi_i \lambda)}{1 + |\sum_i \log(\Psi_i \lambda)|} \right|. \quad (16)$$

if $\mu, \max(r)$, **and** $\mathbf{gap} \leq 0^+$ **then return** λ

else

 Calculate the trust region variable ϕ , as in equation 25

 Calculate $\rho = \phi\mu$, equation 24

end if

Development of the algorithm begins by replacing the constraint $\lambda > 0$ in the primal problem with a log-barrier penalty function, where ρ is a parameter going to 0.

$$\mathcal{P}_\rho \quad \underset{\lambda}{\text{minimize}} \quad \phi(\Psi\lambda) + \overbrace{\rho\phi(\lambda)}^{\text{log-barrier}} + \overbrace{(e^T\lambda - N)}^{\text{penalty function}} \quad \text{subject to } z = \Psi\lambda$$

The third Karush-Kuhn-Tucker condition is changed resulting in

KKT Conditions:

$$\text{for } \rho \downarrow 0, \quad \begin{bmatrix} \Psi^T w + \xi - e_L \\ W\Psi\lambda - e_N \\ \Lambda \Xi e_L - \rho e_L \end{bmatrix} = 0. \quad (17)$$

We now consider solving the perturbed system of nonlinear equations in 17 using a Newton method approach. The Newton method is damped so that the

iterates remain strictly positive. In addition the path following parameter ρ is reduced at each iteration in a manner that attempts to decrease its value at a rate that is approximately the same as the rate of decrease in the error in the nonlinear equation $W\Psi\lambda = e_N$. Specifically, we try to reduce ρ and $\text{norm}(W\Psi\lambda - e_N, \infty)$ at approximately the same rate.

The iterations may thus be described as a sequence of solutions that converge to the solution of equation (3). Each iteration is initialized to satisfy the first KKT condition. In particular, w and ξ are chosen so that the affine equation $\Psi^T w + \xi = e_L$ is satisfied on each iteration. We begin by attempting to set $\lambda = e_L$ and then making the necessary adjustments so that the affine equation is satisfied. The mechanism of the iterative update is described in algorithm 1.

3.2. Newton Step; algorithm 1 steps 1 and 2

The underlying nonlinear function is

$$F(\lambda, w, \xi) = \begin{bmatrix} \Psi^T w + \xi \\ W\Psi\lambda \\ \Lambda \Xi e_L \end{bmatrix}$$

the Jacobian of F is

$$\nabla F(\lambda, w, \xi) = \begin{bmatrix} 0 & \Psi^T & I \\ W\Psi & \text{diag}(\Psi\lambda) & 0 \\ \Xi & 0 & \Lambda \end{bmatrix}$$

and the equation we want to solve is

$$F(\lambda, w, \xi) = [e_L^T, e_N^T, \rho e_L^T]^T$$

The Newton step $(\Delta\lambda, \Delta w, \Delta\xi)$ for this equation at a point (λ, w, ξ) is obtained by solving the linear equation

$$F(\lambda, w, \xi) + \nabla F(\lambda, w, \xi) [\Delta\lambda^T, \Delta w^T, \Delta\xi^T]^T = [e_L^T, e_N^T, \rho e_L^T]^T \quad (18)$$

Using standard reduction techniques we find that

$$\mathbf{H}\Delta w = \begin{cases} W^{-1} [e_N - W\Psi\lambda] - \Psi\Xi^{-1} [\rho e_L - \Xi\Lambda e_L] \\ \quad + \Psi\Lambda\Xi^{-1} [e_L - \Psi^T w - \xi] \\ W^{-1} e_N - \rho\Psi\Xi^{-1} e_L \end{cases} \quad \text{if } e_L = \xi + \Psi^T w \quad (19)$$

$$\mathbf{H} = [W^{-1}\Psi\lambda + \Psi(\Lambda\Xi^{-1})\Psi^T], \text{ is an } N \times N \text{ matrix} \quad (20)$$

$$\Delta\xi = \begin{cases} e_L - \xi - \Psi^T w - \Psi\Delta w \\ -\Psi\Delta w \end{cases} \quad \text{if } e_L = \xi + \Psi^T w \quad (21)$$

$$\Delta\lambda = \rho\Xi^{-1} e_L - \lambda - \lambda\Xi^{-1} \Delta\xi \quad (22)$$

Thus, once we solve for Δw , we can easily compute both $\Delta\xi$ and $\Delta\lambda$. Note that \mathbf{H} is a positive definite symmetric matrix. From here forward, in all our calculations we assume $e_L = \xi + \Psi^T w$.

3.3. Damping; algorithm 1 step 3 and 4

The update of λ , ξ , and w , in equations (14) is damped by the amount calculated in equation (13), with the result that these variables remain strictly positive.

3.4. Relaxation Parameter; algorithm 1 step 5

As optimization progresses the average value of $\lambda_j \xi_j$ goes to zero, this prompts the following definition

$$\text{as } \rho \downarrow 0, \quad \mu = \frac{1}{L} \sum_{j=1}^L \lambda_j \xi_j \rightarrow 0 \quad (23)$$

μ is a boundary parameter of the algorithm. Specifically, the subset of points in G_Ω that will form G_A is found when $\mu \rightarrow 0$.

We attempt to have $\mu \downarrow 0$ at approximately the same rate as the error in the nonlinear function $W\Psi\lambda = e_N$, this is accomplished by setting

$$\rho = \varphi\mu \tag{24}$$

$$\text{where, } \varphi = \begin{cases} 1 & \text{if } \mu < 0^+ < \max(r) \\ \min \left[0.3, (1 - \alpha')^2, (1 - \alpha'')^2, \frac{\max(r) - \mu}{\max(r) + \mu} \right] & \text{otherwise} \end{cases} \tag{25}$$

3.5. Test conditions; algorithm 1 step 5

Convergence of the algorithm to a solution satisfying the exit conditions is implied by theorem 5.24 of [7]. Convergence requires that $\mu \rightarrow 0$, $\max(r) \rightarrow 0$, and **gap** $\rightarrow 0$. $\mu \rightarrow 0$ is discussed above. $\max(r) \rightarrow 0$ states that a potential solution of the optimization has been found: for $i = 1 \dots N$, and Ψ_i the i th row of Ψ , $w_i = 1/\Psi_i\lambda$, see equation (15). **gap** is a measure of the duality gap associated with the solution of primal (λ) and dual (w) problems for the current iteration, see equation (16). The theorems of primal dual theory guarantee that when **gap** goes to zero, the algorithm has converged to the optimal solution.

We are now in a position to discuss details of how NPAG is implemented in a computer program.

Algorithm 2 Grid Expansion. g_A^b is an L -length array of support points. $\varepsilon^0 = 0.2$. New g are sampled from the Euclidian box Ω . Each g_j in g_A is an n -tuple of length $\dim(\Omega)$. New g are the vertices of hypercubes with centers g_j .

$$g_\Omega^{b+1} = g_A^b$$

for $g_j \in g_A^b$ **do**

for $k \in \dim(\Omega)$ **do**

$$\Delta g_j[k] = \varepsilon(b_k - a_k)$$

if $(g_j[k] + \Delta g_j[k] < b_k)$ **then**

$$g^+ = g_j + \Delta g_j[k]$$

$d = 0 \dots$ Verify distance of g^+ to points in g_Ω^{b+1} is greater than θ_d

for $g_l \in g_\Omega^{b+1}$ **do**

$$d = \max(d, \sum_{k=1}^{\dim(\Omega)} \text{abs}(g^+[k] - g_l[k]) / (b_k - a_k))$$

end for

if $d > \theta_d$ **then**

$$g_\Omega^{b+1} = g_\Omega^{b+1} + g^+ \dots \text{Add positive vertex to the expanded grid.}$$

end if

if $(g_j[k] - \Delta g_j[k] > a_k)$ **then**

$$g^- = g_j - \Delta g_j[k]$$

$d = 0 \dots$ Verify distance of g^- to points in g_Ω^{b+1} is greater than θ_d

for $g_l \in g_\Omega^{b+1}$ **do**

$$d = \max(d, \sum_{k=1}^{\dim(\Omega)} \text{abs}(g^-[k] - g_l[k]) / (b_k - a_k))$$

end for

end if

Grid Expansion, continued.

if $d > \theta_d$ **then**

$g_{\Omega}^{b+1} = g_{\Omega}^{b+1} + g^{-}$... Add negative vertex to the expanded grid.

end if

end if

end for

end for

Algorithm 3 NPAG Optimization Loop.

CALL NPAG Optimization Initialization algorithm 4

while $\varepsilon^b > \theta_\varepsilon$ **or** $\|F_1 - F_0\| > \theta_F$ **do**

 CALL IPM(Ψ_Ω^b) find $p(g_\Omega^b)$

 Remove g_j with $p(g_j) < 1 \times 10^{-8}$ from G_Ω “grid condensation” ...

 produce g_A^b

 Determine Ψ_A^b equation 26

 CALL IPM(Ψ_A^b) find $p(g_A^b)$

if γ is being optimized **then**

 CALL OptimizeGamma() algorithm 5

end if

 Calculate fobj^b Test exit conditions

if $\text{abs}(\text{fobj}^b - \text{fobj}^{b-1}) \leq \theta_g$ **and** $\varepsilon^b > \theta_\varepsilon$ **then**

$\varepsilon^{b+1} = \varepsilon^b / 2$

else

$\varepsilon^{b+1} = \varepsilon^b$

end if

if $\varepsilon^{b+1} \leq \theta_\varepsilon$ **then** $F_1 = \text{fobj}^b$

if $\text{abs}(F_1 - F_0) \leq \theta_F$ **then**

 EXIT conditions are met.

else

$F_0 = \text{fobj}^b$; $\varepsilon^{b+1} = 0.2$

end if

end if

NPAG Optimization Loop, continued.

CALL AG(g_A^b) algorithm 2 ... “grid expansion” ... produce g_Ω^{b+1}

Determine Ψ_Ω^{b+1} equation 26

$b = b + 1$ Continue optimization

end while

4. NPAG Implementation

Table 5: Adaptive grid variables.

ε	percent of parameter range used to determine the size of the hypercube surrounding each support point in g_A^b
Δ_{AG}	Difference in likelihood of g_A^B for two successive expansion/contraction sequences
$\theta_F, \theta_\varepsilon, \theta_g, \theta_d$	NPAG control variables

Our approach is to generate a sequence of approximate solutions G_A^b of increasingly greater likelihood, and stopping when evaluation on G_A^{b+1} is negligibly different than evaluation of G_A^b . Each g_A^b is composed of the subset of points of probability $\geq 10^{-8}$, of the larger set g_Ω^b . g_Ω^b is generated from g_A^{b-1} , as described below and in algorithm 2. If G_A^0 is not specified by the user, g_Ω^1 is a Faure sequence of points that uniformly and independently fill Ω [28, 26, 27].

$$G_\Omega^b \xrightarrow{\text{condensation}} G_A^b \xrightarrow{\text{expansion}} G_\Omega^{b+1} \xrightarrow{\text{condensation}} G_A^{b+1} \dots \xrightarrow{\text{condensation}} G_A^B$$

Algorithm 4 NPAG Optimization Initialization.

The initial grid G_A^0 , is input by the user or is a pseudo-random sequence generated using ACM Algorithms 647 and 659 [26, 27]. Set the initial expanded grid, $G_\Omega^1 = G_A^0$, $\gamma = 1$, and $\Delta\gamma = 0.1$

for $g_j \in G_\Omega^1$ **do**

for $y_i \in Y$ **do**

 Determine $\Psi_{i,j}^1$

end for

end for

Set F_0 **and** $\text{fobj}^0 = -1 \times 10^{30}$

Set the grid tolerances, $\theta_d = \theta_g = \theta_\varepsilon = 0.0001$ **and** $\varepsilon^b = 0.2$ **and** $\theta_F = 0.01$

G_A^B is found when the distance between the support points of two successive iterates is less than 0.01% of the range of each dimension of Ω . The sequence above is generated a minimum of two times. Sequences after the first are initialized using the G_A^B of the previous sequence. Global optimum is assumed if evaluation of G_A^B using equation 3 of two successive sequences is negligibly different ($\Delta_{AG} \leq \theta_F$; the default value for θ_F is 0.01). A flowchart of the NPAG program is drawn in figure 3, and more detail is given in Algorithm 3. If the exit conditions are not met prior to a set maximum number of loops, the program will exit after writing the last calculated $\{g_j, \lambda_j\}$ to file.

4.1. Calculation of Ψ

The PK model f , the observations Y , and the noise model, are used to compute the matrix Ψ , which tabulates the likelihood of calculating the i th (row) patient's observations using the j th (column) grid point. Consider equation 1 and assume

that $\Sigma_i = \text{diag}(\sigma_{i,1}^2, \dots, \sigma_{i,M}^2)$. Write

$$y_i = (y_{i,1}, \dots, y_{i,M}), \quad \text{where, } y_{i,m} = f_m(t_m, g_j) + v_{i,m}, \quad \text{and } v_{i,m} \sim N(0, \sigma(f_m(t_m, g_j))^2)$$

It follows that

$$\Psi_{i,j} \triangleq p(y_i | g_j) = \frac{\exp\left(-\frac{1}{2} \sum_{m=1}^M \left(\frac{y_{i,m} - f_m(t_m, g_j)}{\sigma(f_m(t_m, g_j))}\right)^2\right)}{\prod_{m=1}^M \sqrt{2\pi} \sigma(f_m(t_m, g_j))} \quad (26)$$

Our implementation of NPAG approximates σ using a polynomial of the measured concentration. This approximation is sometimes useful for ensuring computational stability and is discussed in section 5.

4.2. Grid expansion

The algorithm begins with an expanded grid, g_Ω^1 , at each subsequent step, we use the current solution supports g_A^b as a base from which to determine a new expanded grid, g_Ω^{b+1} . g_Ω^{b+1} is formed by adding two candidate support points in each dimension of Ω for each support point in g_A^b . The candidate supports are the vertices of a hypercube centered on each $g_j \in g_A^b$ and with segments of length $2\varepsilon(b_k - a_k)$. ε is initially 0.2. As the algorithm generates better solutions to equation 3, the size of the hypercube shrinks, resulting in g_Ω^{b+1} representing a decreasing volume of local space about g_A^b . Since algorithm 1 is employed at each step of NPAG global optimum on the given g_Ω^b and g_A^b is guaranteed. This procedure is described in algorithm 2.

4.3. Support Point Condensation

As shown in figure 3 and algorithm 3, each loop of the algorithm requires two calls to the interior point method. After the first call, the expanded grid is

condensed. Prior to the second call, Householder transformation on a normalized Ψ is used to delete all low probability grid points from consideration. The second call to the interior point method adjusts λ for those remaining grid points. This adjusted λ for the condensed grid is used to form the discrete PDF on Ω output by NPAG.

4.4. Interior Point Method Initialization

The IPM algorithm is initialized to the variable Q_j . This has the effect that initial likelihoods for all g_j are set to a maximal value; optimization has the overall effect of reducing likelihoods of each point until optimum values for all g_j are reached.

$$\begin{aligned} \text{for } j = 1, \dots, L, \quad q_j &= \sum_{i=1}^N \frac{\Psi_{i,j}}{\sum_{j=1}^L \Psi_{i,j}}; \quad Q_j = 2 \max(q_j). \\ \lambda_j^0 &= Q_j \\ \xi_j^0 &= \frac{Q_j - q_j}{Q_j} \end{aligned}$$

$$\begin{aligned} \text{for } i = 1, \dots, N, \quad w_i^0 &= \left(Q_j \sum_{j=1}^L \Psi_{i,j} \right)^{-1} \\ \varphi^0 &= 0 \\ \mathbf{rhs}^0 &= Q_j \Psi e, \quad \text{an } N\text{-length vector.} \end{aligned}$$

Initially, \mathbf{H} is set to the $N \times N$ matrix

$$\mathbf{H}^0(i, k) = \begin{cases} Q_j^2 \sum_{j=1}^L \frac{1}{(Q_j - q_j)} \Psi_{i,j} \Psi_{k,j} & \text{if } i \neq k \\ Q_j^2 \sum_{j=1}^L \left(\frac{1}{(Q_j - q_j)} \Psi_{i,j} + \sum_{j=1}^L \Psi_{i,j} \right) \Psi_{j,j} & \text{if } i = k \end{cases}$$

5. Discussion

NLME modeling is a general hierarchical modeling methodology for repeated measurement data analysis [19, 20]. As a framework, NLME modeling allows inclusion of assumptions regarding the underlying structural model and form of the PK parameter distributions. Mallet [4] proved the optimal form of the population model is a discrete distribution with number of support points at most equal to the number of modeled subjects. Schumitzky [5] then built on the work of Baum et al. [29], Lindsay [3], and Mallet [4], to formulate a discrete EM approach for PK population analysis—the NPEM algorithm, rendering the calculation more approachable, albeit still somewhat time consuming. Leary et al. [6] then introduced an adaptive grid search algorithm, that in conjunction with an IPM (Burke [8] and Bell [9]) vastly improved optimization speed—the NPAG algorithm discussed here.

NPAG is available as a computer program embedded in a package with I/O routines written for population PK modeling [10, Pmetrics]. NPAG is also useful for modeling of systems outside the realm of pharmacokinetics. However, development of algorithms from principles of convex optimization is not trivial. We therefore present here the specific programmed equations as a way of bridging the practical use of NPAG presented in [10] and the theory presented in [7] to make the algorithm more generally accessible.

To complete the description of NPAG we need to model the assay noise. Consider equation 1 and assume that $\Sigma_i = \text{diag}(\sigma_{i,1}^2, \sigma_{i,2}^2, \dots, \sigma_{i,M}^2)$. In the Pmetrics implementation of NPAG, it is further assumed that each $\sigma_{i,m}$ is of the form

$$\sigma'_{i,m} = c_0 + c_1 f_m(t_{i,m}, \beta_i) + c_2 f_m(t_{i,m}, \beta_i)^2 + c_3 f_m(t_{i,m}, \beta_i)^3$$

where one of four different models to represent drug assay noise standard devia-

tion are allowed. Each measurement is then assumed to be $\sim \mathcal{N}(y_{im}, \sigma)$, where

$$\sigma = \begin{cases} \sigma' & \text{assay error polynomial only} \\ \gamma\sigma' & \text{multiplicative error} \\ \sqrt{(\sigma')^2 + \gamma^2} & \text{additive error} \\ \gamma & \text{optimized, constant level of error} \end{cases} \quad (27)$$

If $c_0 = 0$, then $\sigma'_{i,m}$ can become very small for certain values of β_i , which can cause numerical problems. To avoid these problems we assume that $\sigma'_{i,m}$ is *known* and given by

$$\sigma'_{i,m} = c_0 + c_1 y_{i,m} + c_2 y_{i,m}^2 + c_3 y_{i,m}^3 \quad (28)$$

In the above noise models, the assay polynomial coefficients must be put into the algorithm. ² Optimization of γ is explained in algorithm 5.

There are a number of ways to produce the support points and probabilities of a discrete distribution. However, care should be taken when choosing the particular method to use. For example, Leary and Chittenden [18] demonstrated that maximum likelihood optimization over support points with fixed placement based on the MAP Bayesian estimates of a parametric method results in a discrete distribution that inherits the limited exploratory characteristics of the parametric posthoc points. The limitation depends on the particular form of the parametric approximation. In the case of FO or FOCE estimates, which rely on an assumption of normality, shrinkage (Savic and Karlsson [17]) is guaranteed. This is the primary reason for results such as Aldaz et al. [30], where population description of gemcitabine using a 1-compartment model is more variable (and therefore more likely

²A program for estimating the assay error polynomial from patient observations or from assay validation data supplied by the analytic laboratory is available at www.lapk.org.

Algorithm 5 Gamma Optimization. Initially, $\gamma = 1$, and $\Delta\gamma = 0.1$

if γ is being optimized **then**
 for $\gamma = \gamma, \gamma(1 + \Delta\gamma)$, and $\gamma/(1 + \Delta\gamma)$ **do**
 CALL IPM(Ψ_A^b) and calculate fobj
 end for
 Choose best γ
 if γ is adjusted **then**
 $\Delta\gamma = 2 * \Delta\gamma$
 else
 $\Delta\gamma = \max(\Delta\gamma/2, 0.01)$
 end if
end if

to be of clinical value) using NPAG than for a parametric model. An excellent operational definition of NLME modeling using nonparametric maximum likelihood optimization is: A full NPML estimation procedure requires optimization over both probabilities *and* support point positions [18]. To emphasize, fixing placement of support points in a discrete distribution (created with an assumption of an underlying distribution, e.g. normal) induces a parametric assumption onto the population distribution—such a method is *not* nonparametric regardless of what equations are used to determine the likelihoods of the fixed points. However, the optimal solution only occurs if the objective is optimized over both probability and support points.

Sources of variability in drug therapy may be classified as those that are instantaneous, or those that have a cumulative effect over time (process noise). For example, assay error is often approximated by an instantaneous function, whereas

dose amount errors and timing errors in dose or observation affect the reliability of future measurements. Since NPAG only accounts for instantaneous variability (all sources of which may be lumped into the assay error as a product or additive component in equation 27), NPAG is susceptible to modeling error under several conditions that are of clinical importance. Many hospitals do not adhere to recording procedures of time of IV administration or blood draw that enforce reporting time accurately. Patients may have changing PK parameter values during the course of treatment. Patients may not accurately report self-administration of drugs. Each of these situations requires a separate model to describe. We note that the above cautions are true for all PK population modeling algorithms. First, an accurate model must take into account as much as is known apriori regarding the expected uncertainty during application, and second, therapeutic drug monitoring should be used to verify that concentrations are near expectation, and finally, individualization of drug treatment requires adjusting the model to the particular patient. These procedures are required to achieve the desired responses most precisely, regardless of the algorithm used to generate the population PK model.³

Comparing performance on real patient data of the NONP algorithm of NONMEM[©] to the NPAG algorithm of Pmetrics is beyond the purpose of this paper. Several authors have made such a comparison. Each comparison has favored NPAG (see [31] theoretical drug; [32] Cefuroxime Axetil; [33] Ceftazidime; [34] Gabapentin; [30] Gemcitabine; [35] Mycophenolic Acid.)

³The case of individual patients with changing PK parameters in the clinical setting has been modeled by [24]

6. Acknowledgements

We thank the many programmers and staff of LAPK, who have aided users of NPAG with their various needs over the past 10 years. The research and analysis presented in this paper was supported in part by NIH/NLMG-R01LM05401, NIH/NCRR-R01RR11526, NIH/NLM-RO1 LM05401, NIH/NIGMS-R01 GM65619, NIH/NIGMS-R01 GM68968, NIH/NIBIB-RO1 EB005803, NIH/NIGMS-R01GM068968, and NIH/NICHHD-R01HD070886. Linear algebra routines supplied by [36] were used, especially for solution of equation (10). JB was supported in part by NSF/DMS-0505712.

7. Conflict of interest statement

The authors have no potential conflict of interest that might bias their report of the NPAG algorithm.

References

- [1] M. Rowland, L. Sheiner, J.-L. Steimer (Eds.), *Variability in Drug Therapy: Description, Estimation, and Control*, Raven Press, 1185 Avenue of the Americas, New York, New York, USA, 1985.
- [2] J. Rodman, D. D'Argenio, C. Peck, *Applied Pharmacokinetics and Pharmacodynamics: Principles of Therapeutic Drug Monitoring*, fourth ed., Lippincott Williams and Wilkins, 2006, pp. 40–59.
- [3] B. G. Lindsay, The geometry of mixture likelihoods: A general theory, *Ann. Statist.* 11 (1983) 86–94.

- [4] A. Mallet, A maximum likelihood estimation method for random coefficient regression models, *Biometrika* 73 (1986) 645–656.
- [5] A. Schumitzky, Nonparametric EM algorithms for estimating prior distributions, *Applied Mathematics and Computation* 45 (1991) 143–157.
- [6] R. Leary, R. Jelliffe, A. Schumitzky, M. V. Guilder, An adaptive grid non-parametric approach to pharmacokinetic and dynamic (PK/PD) population models, 14th IEEE Symposium on Computer-Based Medical Systems (CBMS'01) 0 (2001) 0389.
- [7] Y. Baek, An Interior Point Approach to Constrained Nonparametric Mixture Models, PhD dissertation, Letters and Science, 2006.
- [8] J. Burke, Interior point method matlab script, <http://www.math.washington.edu/~burke/>, 2013.
- [9] B. Bell, Brad's opens source packages, <http://www.seanet.com/~bradb主ell/packages.htm>, 2013.
- [10] M. Neely, M. van Guilder, W. Yamada, A. Schumitzky, R. Jelliffe, Accurate detection of outliers and subpopulations with pmetrics: a non-parametric and parametric pharmacometric package for R, *Therapeutic Drug Monitoring* 34 (2012) 467–476.
- [11] R. Koenker, I. Mizera, Convex optimization, shape constraints, compound decisions, and empirical bayes rules, <http://www.econ.uiuc.edu/~roger/research/ebayes/brown.pdf>, 2013.

- [12] R. Jelliffe, Goal-oriented, model-based drug regimens: Setting individualized goals for each patient, *Therapeutic Drug Monitoring* 22 (2000) 320–325.
- [13] J. Kiefer, J. Wofowitz, Consistency of the maximum likelihood estimator in the presence of infinitely many incidental parameters, *Ann. Math. Statist.* 27 (1956) 887–906.
- [14] L. Sheiner, S. Beal, NONMEM Users Guide, Technical Report, University of California, San Francisco, San Francisco, CA, USA, 1992.
- [15] U.S. Department of Health and Human Services, Guidance for Industry Population Pharmacokinetics, Technical Report, United States Department of Health and Human Services, Food and Drug Administration, 1999. This report is specific to using NONMEM. However, page 13, section VII.C.Outliers, points toward the work of [4] and [37] as better approaches.
- [16] F. Mentré, A. Mallet, Handling covariates in population pharmacokinetics, *international Journal of Biomedical Computing* 36 (1994) 25–33.
- [17] R. M. Savic, M. O. Karlsson, Importance of shrinkage in empirical bayes estimates for diagnostics and estimation: Problems and solutions, in: Abstracts of the Annual Meeting of the Population Approach Group in Europe, PAGE2007, 2007.
- [18] R. Leary, J. Chittenden, A nonparametric analogue to posthoc estimates for exploratory data analysis, in: Abstracts of the Annual Meeting of the Population Approach Group in Europe, PAGE2008, 2008, p. 17.

- [19] M. Davidian, D. M. Giltinan, *Nonlinear Models for Repeated Measurement Data*, Chapman and Hall/CRC Press, 1995.
- [20] M. Davidian, D. M. Giltinan, Nonlinear models for repeated measurement data: An overview and update, *Journal of Agricultural, Biological, and Environmental Statistics* 8 (2003) 387–419.
- [21] R. Jelliffe, D. Bayard, M. Milman, M. V. Guilder, A. Schumitzky, Achieving target goals most precisely using nonparametric compartmental models and ‘multiple model’ design of dosage regimens, *Therapeutic Drug Monitoring* 22 (2000) 346–353.
- [22] European Medicines Agency, *Guideline on Reporting the Results of Population Pharmacokinetic Analyses*, Technical Report, European Medicines Agency, 7 Westferry Circus, Canary Wharf, London, E14 4HB, UK, 2007. This report is specific to the use of NONMEM.
- [23] C. Carathéodory, über den variabilitätsbereich der fourierschen konstanten von positiven harmonischen funktionen, *Rend. Circ. Mat. Palermo* 32 (1911) 193–217.
- [24] D. Bayard, R. Jelliffe, A bayesian approach to tracking patients having changing pharmacokinetic parameters, *Journal of Pharmacokinetics and Pharmacodynamics* 31 (2004) 75–107.
- [25] S. Boyd, L. Vandenberghe, *Convex Optimization*, Cambridge University-Press, 2004.
- [26] P. Bratley, B. L. Fox, *Algorithm 659: Implementing Sobol’s quasirandom*

- sequence generator, *ACM Transactions on Mathematical Software* 14 (1988) 88–100. [Http://www.netlib.org/toms/659](http://www.netlib.org/toms/659).
- [27] B. L. Fox, Algorithm 647: Implementation and relative efficiency of quasirandom sequence generators, *ACM Transactions on Mathematical Software* 12 (1986) 362–376. [Http://www.netlib.org/toms/647](http://www.netlib.org/toms/647).
- [28] H. Faure, Discrépance de suites associées á un système de numération (en dimension s), *Acta Arithmetica* 41 (1982) 337–351.
- [29] L. E. Baum, T. Petrie, G. Soules, N. Weiss, A maximization technique occurring in the statistical analysis of probabilistic functions of markov chains, *Ann. Math. Statist.* 41 (1970) 164–171.
- [30] A. Aldaz, O. Sayar, L. Zufia, A. Viudez, J. Rifon, Y. Nieto, Comparison between NONMEM and NPAG for gemcitabine modelling, in: Abstracts of the Annual Meeting of the Population Approach Group in Europe, PAGE2007, 2007.
- [31] J. Antic, C. Laffont, D. Chafaï, D. Concordet, Comparison of nonparametric methods in nonlinear mixed effects models, *Computational Statistics and Data Analysis* 53 (2009) 642–656.
- [32] J. Bulitta, C. Landersdorfer, M. K. and. U. Holzgrabe, F. Sorgel, New semi-physiological absorption model to assess the pharmacodynamic profile of cefuroxime axetil using nonparametric and parametric population pharmacokinetics, *Antimicrobial Agents and Chemotherapy* 53 (2009) 3462–3471.
- [33] J. Bulitta, C. Landersdorfer, S. Hüttner, G. Drusano, M. K. and. U. Holzgrabe, U. Stephan, F. Sorgel, Population pharmacokinetic comparison and

- pharmacodynamic breakpoints of ceftazidime in cystic fibrosis patients and healthy volunteers, *Antimicrobial Agents and Chemotherapy* 54 (2010) 1275–1282.
- [34] K. Carlsson, M. van de Schootbrugge, H. Eriksen, E. Moberg, M. Karsson, N. Hoem, A population pharmacokinetic model of gabapentin developed in nonparametric adaptive grid and nonlinear mixed effects modeling, *Therapeutic Drug Monitoring* 31(1) (2009) 86–94.
- [35] A. Prémaud, L. T. Weber, B. Tönshoff, V. Armstrong, M. Oellerich, S. Urien, P. Marquet, A. Rousseau, Population pharmacokinetics of mycophenolic acid in pediatric renal transplant patients using parametric and nonparametric approaches, *Pharmacological Research* 63 (2011) 216–224.
- [36] E. Anderson, Z. Bai, C. Bischof, S. Blackford, J. Demmel, J. Dongarra, J. Du Croz, A. Greenbaum, S. Hammarling, A. McKenney, D. Sorensen, *LAPACK Users' Guide*, third ed., Society for Industrial and Applied Mathematics, Philadelphia, PA, 1999.
- [37] J. Wakefield, The bayesian analysis of population pharmacokinetic models, *J Am Stat Assoc* 91 (1996) 62–75.