

Maximum Likelihood Estimators of Brain White Matter Microstructure

Olivier Commowick, Aymeric Stamm, Simone Vantini, Simon K. Warfield

▶ **To cite this version:**

Olivier Commowick, Aymeric Stamm, Simone Vantini, Simon K. Warfield. Maximum Likelihood Estimators of Brain White Matter Microstructure. 48th Scientific Meeting of the Italian Statistical Society, Jun 2016, Salerno, Italy. inserm-01347200

HAL Id: inserm-01347200

https://inserm.hal.science/inserm-01347200

Submitted on 20 Jul 2016

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Maximum Likelihood Estimators of Brain White Matter Microstructure

Stimatori di massima verosimiglianza per la microstruttura della materia bianca cerebrale

Olivier Commowick[1], Aymeric Stamm[2,3], Simone Vantini[2], and Simon K. Warfield[3]

Abstract The microstructure of the brain white matter is not visible to the naked eye but would be of invaluable help to the clinician in the diagnosis and treatment of many brain pathologies. Diffusion MRI is an in-vivo non invasive imaging technique that probes the cyto-architecture of the white matter through the diffusion of water. However, diffusion MRI is limited in resolution, which makes forward models of the diffusion at the voxel level rather complex. In this paper, we provide a statistical framework for recovering the maximum-likelihood estimators of the parameters of mixture models of the diffusion. We calibrate different methods on simulated data to guarantee convergence to the maximum likelihood and show that profile likelihood maximization using variable projection together with a Levenberg-Marquardt algorithm with analytic Jacobian is the most efficient method to obtain the MLE.

Abstract *La microstruttura della materia bianca del cervello umano è, da un lato, non visibile ad occhio nudo e, dall'altro, portatrice di un indubbio valore clinico sia in termini di diagnosi che di trattamento delle patologie cerebrali. La risonanza magnetica di diffusione (dMRI) è una pratica medica per l'esplorazione in-vivo e non invasiva della cito-architettura della materia bianca basata sulla diffusione dell'acqua. Tuttavia, la risoluzione spaziale limitata della dMRI rende necessario l'utilizzo di modelli di diffusione molto complessi a livello del singolo voxel. Questo lavoro si focalizza sulla stima di massima verosimiglianza dei parametri dei modelli di diffusione di tipo mistura utilizzati in questo contesto. I vari metodi sono confrontati sfruttando dati simulati per avere garanzia di convergenza al massimo della*

[1]Olivier Commowick
VISAGES: INSERM U746 - CNRS UMR6074 - INRIA - Univ. of Rennes I, France
e-mail: olivier.commowick@inria.fr

[2]Aymeric Stamm and Simone Vantini
MOX, Department of Mathematics, Politecnico di Milano, Milan, Italy
e-mail: aymeric.stamm@polimi.it, simone.vantini@polimi.it

[3]Aymeric Stamm and Simon K. Warfield
CRL, Boston Children's Hospital, Harvard Medical School, Boston, MA, USA
e-mail: aymeric.stamm@childrens.harvard.edu, simon.warfield@childrens.harvard.edu

verosimiglianza. Il confronto mette in luce come la massima efficienza venga ottenuta mediante la massimizzazione della profile-likelihood tramite il metodo della proiezione di variabili accopiato all'uso dell'algoritmo di Levenberg-Marquardt con Jacobiano analitico.

Key words: maximum likelihood, variable projection, diffusion MRI, compartment models, brain, white matter.

1 Introduction

White matter in the human brain is the seat of neuron projections that are responsible for transmitting nerve impulses that govern the functions of our body. These projections, called axons, are cylindrically shaped and connect different functional areas of the brain together. They are surrounded by spherically shaped glial cells whose function is to support neuronal activity. Both have impermeable membranes in a healthy subject. While axons and glial cells are invisible to the naked eye, there is an urgent unmet need to visualize them in-vivo. In effect, we can think of the example of a neurosurgeon who needs to remove a tumor: he needs to know if critical connexions are on his access path to the tumor or within the resection area so that no irreversible damage is done to functions such as breathing or moving.

Diffusion MRI provides us with the means to observe these axons and cells by the observation that water inside axons or glial cells is subject to restricted diffusion as it bounces on and off the membranes. Hence, inferring the probability distribution of 3D molecular displacements due to diffusion provides knowledge about tissues themselves. Nevertheless, spatial resolution is currently limited to $2 \times 2 \times 2\text{mm}^3$, while individual cell diameters are of the order of the micron. Many cells and various axon bundles with different orientations will then populate a white matter voxel and the underlying diffusion process that needs to be estimated can be quite complex.

At the voxel level, the diffusion is naturally modeled using so-called *multi-compartment models* [6, 9], known by statisticians as mixture models. A *compartment* is defined as a sub-volume of the voxel in which random molecular motion is governed by the same underlying probability distribution, with the assumption of no exchange between compartments. The focus of this work is not to discuss the validity and/or performances of different compartmental diffusion models that have been proposed over the years, but rather to concentrate on a generic and robust framework for the estimation of those mixture models from the acquired noisy data. We start upfront by assuming compartmental diffusion to be a zero-mean Gaussian process. The resulting multi-compartment model is often dubbed *multi-tensor model*, since each compartment is identified by a diffusion tensor akin to a 3D covariance matrix.

The genesis of this work relies on the observation that, in the computational MR community, focus was on designing the best possible model but little attention has been paid on how well these models are estimated. In effect, Gaussian mixture model estimation is an extremely intricate problem to solve while state-of-the-art ap-

proaches to their estimation consist in attempting a minimization of the least squares criterion between observations and signals predicted by the multi-tensor model using a derivative-free algorithm which features computation times of the order of the week on super-computers with no convergence guarantee [9]. There has been attempts at overcoming this issue using a Levenberg-Marquardt algorithm but, so far, it has been limited to two compartments since this algorithm cannot handle inequality constraints on the parameters. In addition, convergence is barely discussed and statistical properties of underlying estimators are often disregarded or unknown.

This is however critical both for fair model comparison and, even more importantly, in providing accurate and reproducible results to the clinician. Our purpose is thus to provide a systematic statistical framework for finding the maximum-likelihood estimators (MLE) of the parameters defining the multi-tensor model in a minimum amount of computation time with reasonable convergence towards the global maximum. The paper is structured as follows: in Section 2, we quickly review the voxelwise relationship between the signal measured in dMRI and the underlying diffusion process and we recall the formulation of the multi-tensor model. In Section 3, we introduce our strategies to design maximum likelihood estimators for the parameters of the diffusion mixture models. Finally, in Section 4, we calibrate different estimation strategies on simulated data to ensure convergence to the MLE and show that using a Levenberg-Marquardt algorithm with analytic Jacobian to maximize the profile likelihood using variable projection yields the most efficient strategies to get the MLE of diffusion mixture models.

2 Diffusion MRI and the Multi-Tensor Model

In diffusion MRI, the subject's head lies in a tube immersed in a strong magnetic field. Protons of water in his brain are excited by radio-frequencies, which generate a net magnetization measurable by an array of coils as a current. We shall denote by μ_0 the true amplitude of such a signal. Next, a collection of N magnetic field spatial gradients, with different magnitudes and directions, are applied, resulting in a decay of the net magnetization due to the random diffusion of water. If we assume that the probability distribution that governs diffusing water motion is a Gaussian mixture with zero mean, the true net magnetization decay S_i/μ_0 resulting from the application of a magnetic field spatial gradient $(b_i, \mathbf{g}_i)$ reads:

$$\frac{\widehat{y_i}(p_1, \ldots, p_C, D_0, \ldots, D_C; b_i, \mathbf{g}_i)}{\mu_0} = \left(1 - \sum_{j=1}^{C} p_j\right) e^{-b_i \mathbf{g}_i^\top D_0 \mathbf{g}_i} + \sum_{j=1}^{C} p_j e^{-b_i \mathbf{g}_i^\top D_j \mathbf{g}_i}, \quad (1)$$

where $\mathbf{g_i}$ is the direction of the gradient, b_i is a positive value proportional to the gradient magnitude, p_j's are the mixture weights and D_j's are the mixture covariance matrices, often termed *tensors*, hence the full name multi-tensor model.

Microscopy studies have shown that axons tend to regroup into dense bundles with a common orientation. We shall refer to these bundles of axons as *fascicles*. The model assumes that, in a voxel, water can be subject to 4 types of diffusion:

- Free Water (FW): when water is not trapped within cells and away from cell membranes, it is subject to free isotropic diffusion. The multi-tensor model includes FW in a single mixture component featuring isotropic covariance with eigenvalue equal to $d_{\text{FW}} = 0.003\text{mm}^2/\text{s}$, which is the diffusivity of FW at 37°C.
- Stationary Water (SW). The multi-tensor includes one mixture component with null covariance matrix to account for water that does not diffuse at all.
- Isotropically Restricted Water (IRW): water inside glial cells can diffuse in all directions evenly but is restricted by cell membranes. The multi-tensor model includes IRW in a single mixture component featuring isotropic covariance with eigenvalue equal to $d_{\text{IRW}} = 0.001\text{mm}^2/\text{s}$ as determined in [7].
- Anisotropically Restricted Water (ARW): water trapped inside fascicles will diffuse mainly along the fascicle orientation and motion is restricted by axon membranes. The multi-tensor model accommodates ARW in C different mixture components, to account for multiple fascicles that cross in the voxel, in which covariance matrices are expected to be anisotropic, i.e. with a principal eigenvalue much larger than the other two.

In summary, the multi-tensor model considered here has $7C+2$ free parameters to be estimated: the 6 covariance entries, the mixture weight for each of the C fascicle components and the 2 mixture weights of two out of the three isotropic components (the last one being obtained by the constraint that weights should sum up to one). In addition, the net magnetization μ_0 is a nuisance parameter that needs to be estimated. The number C of fascicle components usually varies between 0 and 3 depending on the location of the voxel in the brain [10]. The forward model thus reads:

$$\frac{\widehat{y_i}(p_{\text{SW}}, p_{\text{IRW}}, p_1, \ldots, p_C, D_1, \ldots, D_C; b_i, \mathbf{g}_i)}{\mu_0} = \left(1 - p_{\text{SW}} - p_{\text{IRW}} - \sum_{j=1}^{C} p_j\right) e^{-b_i d_{\text{FW}}} + p_{\text{SW}} e^{-b_i d_{\text{SW}}} + p_{\text{IRW}} e^{-b_i d_{\text{IRW}}} + \sum_{j=1}^{C} p_j e^{-b_i \mathbf{g}_i^\top D_j \mathbf{g}_i}. \quad (2)$$

In the next section, we shall present three possible strategies to perform maximum likelihood estimation for the parameters of such a forward model.

3 Maximum Likelihood Estimators

The problem of defining maximum likelihood estimators (MLE) is intrinsically related to the nature of the measurement noise. In MRI, the noise in the raw measurements made by the individual surface coils follows a complex Gaussian distribution, and can, in principle, be effectively combined into a composite real-valued Gaussian

random measure ready for diffusion modeling [11]. When such a reconstruction is sub-optimal, it can be shown that the measured diffusion signal follows a non-central χ-distribution [1], which is fairly well approximated by a Gaussian one, provided that signal-to-noise ratio is high. Hence, we focus on maximizing the likelihood of model parameters given Gaussian-corrupted input diffusion signals with homogeneous variance. For clarity and conciseness, given that the estimation problem is solved independently in each voxel of the 3D image of the brain, we shall focus on the single voxel and thus drop any index that refers to it.

Assume that we collected N diffusion signals $y_1, \ldots, y_N$ by applying N non-collinear magnetic field spatial gradients $(b_1, \mathbf{g}_1), \ldots, (b_N, \mathbf{g}_N)$. First, observe that we can write Eq. (2) in matrix form as:

$$
\begin{aligned}
&\widehat{\mathbf{y}} = \mu_0 \left[\mathbf{a}_0 + \Phi(D_1, \ldots, D_C)\mathbf{w}\right] \in \mathbb{R}^N, \quad \text{where:} \\
&a_{i0} = e^{-b_i d_{\text{FW}}},\ a_{i1} = e^{-b_i d_{\text{SW}}},\ a_{i2} = e^{-b_i d_{\text{IRW}}},\ a_{ij}(D_j) = e^{-b_i \mathbf{g}_i^\top D_j \mathbf{g}_i} \text{ for } j \geq 3 \\
&\mathbf{w} = (p_{\text{SW}}, p_{\text{IRW}}, p_1, \ldots, p_C)^\top \in \mathbb{R}^{C+2}, \\
&\Phi(D_1, \ldots, D_C) \in \mathbb{R}^{N \times (C+2)} \text{ s.t. } \phi_{i1}(D_1, \ldots, D_C) = a_{i1} - a_{i0}, \\
&\phi_{i2}(D_1, \ldots, D_C) = a_{i2} - a_{i0}, \text{ and } \phi_{ij}(D_1, \ldots, D_C) = a_{ij}(D_j) - a_{i0}, \text{ for } j \geq 3.
\end{aligned}
$$

Hence, the likelihood of the model parameters under the assumption of white noise with precision parameter $\tau^2 = 1/\sigma^2$ reads:

$$
L(\mu_0, \tau^2, \mathbf{w}, \{D_j\}; \mathbf{y}) = \left(\frac{\tau^2}{2\pi}\right)^{N/2} \exp\left\{-\frac{\tau^2}{2} \left\|\mathbf{y} - \mu_0 \left[\mathbf{a}_0 + \Phi(D_1, \ldots, D_C)\mathbf{w}\right]\right\|^2\right\}. \quad (3)
$$

Section 3.1 outlines the 3 different strategies to find the MLEs of the model parameters maximizing Eq. (3). Section 3.2 focuses on optimization algorithms required in any strategy for finding the MLE of at least a subgroup of parameters.

3.1 Strategies for Maximum Likelihood Estimation

Marginal Likelihood Maximization. This strategy pertains to integrating out of the likelihood any parameter considered as a nuisance parameter. In dMRI, we are very much interested in the mixture weights $\mathbf{w}$ and the tensors D_j's. However, μ_0 and τ^2 are irrelevant parameters, although we need to estimate them accurately as well. Hence, we can integrate the likelihood given in Eq. (3) over μ_0 and τ^2 to obtain a marginal likelihood that depends only on the parameters of interest. The latter ones are then found by maximizing the marginal log-likelihood using non-linear optimization algorithms. We denote this strategy as MARGINAL.

Profile Likelihood Maximization. This strategy pertains to profiling out some of the model parameters in terms of the others. For the current problem, we observe that for a set of fixed diffusion parameters $\mathbf{w}$ and D_j's, we can find analytically the values of $\widehat{\mu_0}(\mathbf{w}, \{D_j\})$ and $\widehat{\tau^2}(\mathbf{w}, \{D_j\})$ that maximize the log-likelihood. We

can then substitute these expressions in the original log-likelihood function and maximize the resulting profile log-likelihood using optimization algorithms to solve for the diffusion parameters. We denote this strategy as PROFILE.

Profile Likelihood by Variable Projection. This strategy is also based on profile likelihood maximization. It uses the fact that the log-likelihood that we want to maximize is actually linear in the mixture weights. Hence one can start by profiling out the mixture weights using linear least-square methods and subsequently profile out μ_0 and τ^2. Consequently, we only have to resort to an optimization algorithm to find the tensors D_j's. This is known in the optimization literature as variable projection [3]. We denote this strategy as VARPRO.

3.2 Optimization Algorithms

In each of the three strategies previously described, at least the MLE of the tensors D_j's requires the use of algorithms for non-linear optimization. All optimized parameters are bounded: angles and eigenvalues defining the tensors are bounded by biological/physical constraints and mixture weights are bounded to $[0, 1]$ in that they represent proportions. In the current work, we compare three different algorithms:

Bounded Optimization BY Quadratic Approximations (BOBYQA) [8]. It is a derivative-free algorithm that solves bound constrained optimization problems. We use the implementation of the NLOpt library [4]. It is one of the two standard algorithms used in the literature for fitting multi-tensor models [9].

Levenberg Marquardt algorithm [5]. This algorithm deals with unconstrained non-linear least square problems only. We use the implementation of the Insight ToolKit (ITK)[1] after mapping the original bounded parameters onto the real line. It is gradient-based but the Jacobian is not mandatory. It can be computed by forward differences if not provided. This algorithm with forward difference approximation of the Jacobian is the second of the two standard algorithms used in the literature of multi-tensor models [6].

Conservative Convex Separable Approximation (CCSA) [12]. It is a gradient-based algorithm that supports bound constraints. We use the implementation from the NLOpt library. It requires the user to provide the Jacobian.

Finally, whenever the mixture weights have to be optimized as well through these algorithms, their sum should not exceed 1. This further inequality constraint is nicely accommodated in the NLOpt library, which offers an Augmented Lagrangian method [2] to cope with inequality constraints. No such method exists for the Levenberg Marquardt algorithm, which will thus only be used with the VARPRO strategy.

[1] http://www.itk.org

4 Simulations & Discussion

We have designed a simulation that allows to calibrate the termination criterion of the various algorithms to ensure reasonable convergence towards the maximum of the log-likelihood function and compare the resulting computation times to provide guidelines to the best strategy for finding ML estimators for the multi-tensor model.

We simulated theoretical diffusion signals out of the forward multi-tensor model using all three isotropic compartments and one fascicle compartment. We used two sets of magnetic field spatial gradients with $N_1 = 65$ (ACQ1) an $N_2 = 288$ (ACQ2) gradients respectively. The net magnetization before applying any gradient was set to $\mu_0 = 3300$. Proportions of the compartments were set to $p_{\text{FW}} = 0.07$, $p_{\text{SW}} = 0.03$, $p_{\text{IRW}} = 0.1$ and $p_{\text{ARW}} = 0.8$. The tensor of the ARW compartment was cigar-shaped with a first eigenvalue much larger w.r.t. the other two. Eigenvalues and orientations vary voxelwise. For each set of gradients, we added Gaussian noise to the ground truth diffusion signals with standard deviation of 8% of the net magnetization μ_0.

Strategy	Algorithm	Time (s)	Avg. Rel. Error on σ^2 (%)	Avg. Rel. Error on μ_0 (%)	Avg. Quad. Error on $\mathbf{w}$ $\left(\times 10^{-2}\right)$
MARGINAL	BOBYQA	152	−13.4714 (±16.4024)	0.2242 (±3.5420)	0.9408 (±1.0392)
MARGINAL	CCSA	363	−13.9121 (±16.3688)	0.2469 (±3.5084)	1.1160 (±1.1788)
PROFILE	BOBYQA	155	−13.4661 (±16.4109)	0.1565 (±3.5497)	0.9402 (±1.0523)
PROFILE	CCSA	361	−13.9135 (±16.3684)	0.1763 (±3.5116)	1.1163 (±1.1808)
VARPRO	BOBYQA	992	−13.9210 (±16.3665)	0.1650 (±3.5138)	1.1309 (±1.1766)
VARPRO	CCSA	224	−13.9203 (±16.3666)	0.1637 (±3.5144)	1.1306 (±1.1775)
VARPRO	LM (approx)	467	−13.9210 (±16.3665)	0.1650 (±3.5137)	1.1310 (±1.1767)
VARPRO	LM (exact)	79	−13.9135 (±16.3668)	0.1695 (±3.5105)	1.1304 (±1.1760)

Table 1 Estimation results with the ACQ1 protocol ($N = 65$).

Tables 1 and 2 show the results of the estimation of the multi-tensor model, where LM (approx) stands for Levenberg Marquardt algorithm with forward difference Jacobian approximation while LM (exact) uses the analytic Jacobian. The tables show the mean and standard deviation of the relative error on the noise variance σ^2, on the μ_0 parameters and the relative quadratic error on the mixture weights $\mathbf{w}$.

Strategy	Algorithm	Time (s)	Avg. Rel. Error on σ^2 (%)	Avg. Rel. Error on μ_0 (%)	Avg. Quad. Error on $\mathbf{w}$ $\left(\times 10^{-2}\right)$
MARGINAL	BOBYQA	268	−2.9309 (±8.2558)	−0.3881 (±1.9206)	0.4741 (±0.4428)
MARGINAL	CCSA	1039	−3.3113 (±8.2423)	−0.3601 (±1.8785)	0.4642 (±0.4842)
PROFILE	BOBYQA	267	−2.9285 (±8.2575)	−0.4075 (±1.9301)	0.4767 (±0.4246)
PROFILE	CCSA	1040	−3.3113 (±8.2424)	−0.3826 (±1.8797)	0.4641 (±0.4853)
VARPRO	BOBYQA	2417	−3.3162 (±8.2416)	−0.3845 (±1.8805)	0.4589 (±0.4637)
VARPRO	CCSA	631	−3.3161 (±8.2416)	−0.3845 (±1.8805)	0.4588 (±0.4639)
VARPRO	LM (approx)	1066	−3.3162 (±8.2416)	−0.3845 (±1.8805)	0.4589 (±0.4638)
VARPRO	LM (exact)	230	−3.3162 (±8.2416)	−0.3844 (±1.8805)	0.4589 (±0.4638)

Table 2 Estimation results with the ACQ2 protocol ($N = 288$).

Having in mind that the log-likelihood is inversely proportional to the MLE of the variance parameter, convergence to the global maximum of the likelihood is reached whenever the average relative error on σ^2 is minimal, i.e., whenever $\widehat{\sigma^2}$ is minimal. That being said, it is clear from Table 2 that the LM (exact) combined with VARPRO is the quickest way to reach the global maximum. Table 1 seems to suggest that this method reaches a value very close but not equal to the global maximum. Nevertheless, given that it provides a 6-fold acceleration w.r.t. to the quickest method that finds the global maximum and that the difference in mixture weight estimates is almost irrelevant, these results yield the same conclusion.

In summary, we highly recommend the use of the variable projection method together with the Levenberg Marquardt algorithm with exact derivative to solve the problem of multi-tensor model estimation. This is very effective in providing the ML estimate of the parameters in a short computation time.

Acknowlegments. This investigation was supported in part by NIH grants R01 NS079788 and R01 DK100404.

References

1. S. Aja-Fernandez, V. Brion, and A. Tristan-Vega. Effective noise estimation and filtering from correlated multiple-coil MR data. *Magn. Reson. Imaging*, 31(2):272–85, 2013.
2. A. R. Conn, N. I. M. Gould, and P. Toint. A Globally Convergent Augmented Lagrangian Algorithm for Optimization with General Constraints and Simple Bounds. *SIAM J. Numer. Anal.*, 28(2):545–572, 1991.
3. G.H. Golub and V. Pereyra. The differentation of pseudo-inverses and nonlinear least squares problems whose variables separate. *SIAM J. Numer. Anal.*, 10(2):413–432, 1973.
4. S.G. Johnson. The NLopt nonlinear-optimization package. http://ab-initio.mit.edu/nlopt.
5. D.W. Marquardt. An Algorithm for Least-Squares Estimation of Nonlinear Parameters. *J. Soc. Ind. Appl. Math.*, 11(2):431–441, 1963.
6. E. Panagiotaki, T. Schneider, B. Siow, M.G. Hall, M.F. Lythgoe, and D.C. Alexander. Compartment models of the diffusion MR signal in brain white matter: a taxonomy and comparison. *Neuroimage*, 59(3):2241–54, 2012.
7. J. Pfeuffer, U. Flögel, W. Dreher, and D. Leibfritz. Restricted diffusion and exchange of intracellular water: theoretical modelling and diffusion time dependence of 1H NMR measurements on perfused glial cells. *NMR Biomed.*, 11(1):19–31, 1998.
8. M.J.D. Powell. The BOBYQA algorithm for bound constrained optimization without derivatives. Technical report, University of Cambridge, 2009.
9. B. Scherrer and S.K. Warfield. Parametric representation of multiple white matter fascicles from cube and sphere diffusion MRI. *PLoS One*, 7(11):e48232, 2012.
10. A. Stamm, O. Commowick, P. Pérez, and C. Barillot. Fast identification of optimal fascicle configurations from standard clinical diffusion MRI using Akaike information criterion. In *IEEE Int. Symp. Biomed. Imaging*, pages 238–241, 2014.
11. A. Stamm, J. Singh, O. Afacan, and S.K. Warfield. Analytic Quantification of Bias and Variance of Coil Sensitivity Profile Estimators for Improved Image Reconstruction in MRI. In *MICCAI*, LNCS, pages 684–691. Springer, 2015.
12. K. Svanberg. A Class of Globally Convergent Optimization Methods Based on Conservative Convex Separable Approximations. *SIAM J. Optim.*, 12(2):555–573, 2002.