Reviews in Computational Chemistry 111

Edited by
Kenny B. Lipkowtiz and Donald B. Boyd
Reviews in Computational Chemistry Volume III
A NOTE TO THE READER
This book has been electronically reproduced from digital information stored at John Wiley & Sons, Inc. We are pleased that the use of this new technology will enable us to keep works of enduring scholarly value in print as long as there is reasonable demand for them. The content of this book is identical to previous printings.

Copyright © 1992 by John Wiley & Sons, Inc. All rights reserved.

Originally published as ISBN 1-56081-619-8

No part of this publication may be reproduced, stored in a retrieval system, or transmitted in any form or by any means, electronic, mechanical, photocopying, recording, scanning or otherwise, except as permitted under Sections 107 and 108 of the 1976 United States Copyright Act, without either the prior written permission of the Publisher, or authorization through payment of the appropriate per-copy fee to the Copyright Clearance Center, 222 Rosewood Drive, Danvers, MA 01923, (978) 750-8400, fax (978) 750-4744. Requests to the Publisher for permission should be addressed to the Permissions Department, John Wiley & Sons, Inc., 605 Third Avenue, New York, NY 10158-0012, (212) 850-6011, fax (212) 850-6008, E-mail PERMREQ@WILEY.COM.

For ordering and customer service, call 1-800-CALL-WILEY.

Library of Congress Cataloging-in-Publication Data

Reviews in computational chemistry / edited by Kenny B. Lipkowitz and Donald B. Boyd
p. cm.
Includes bibliographical references and index.
ISBN 0-471-18853-0
I. Lipkowitz, Kenny B. II. Boyd, Donald B.
QD39.3.E46R5 1993 92-30192
541'.2'2 — dc20

Printed in the United States of America.
10 9 8 7 6 5 4 3 2 1
Computational chemistry, as a discipline, is still growing steadily. Although it has historically been aligned with physical chemistry, computational chemistry has taken on a vibrant life of its own and spread into other domains of chemical research and education. In some disciplines such as organic chemistry, molecular modeling has almost become a part of the regular routine of research. Other disciplines, such as polymer chemistry, have only recently begun to integrate computational approaches to solving problems. At the 1992 Spring National American Chemical Society meeting, 20% of the 636 presentations in the polymer division were classified as modeling and computer simulation. At that same meeting, fully 70% of the exposition workshops involved computational chemistry. One may argue that this is because computers are easy to carry to conventions, but we maintain that there is great interest in computational chemistry because it is a "hot" field. This ACS meeting is not exceptional.

The number of meetings touching on aspects of computational chemistry is truly amazing. For instance, during a single two-month period in 1991, more than 30 conferences, symposia, and workshops were scheduled.* If these meet-

*The following were among the meetings held in May and June 1991: 1st Canadian Symposium on Computational Chemistry (Orford, Quebec); 2nd International Symposium on Chiral Discrimination (Rome, Italy); 3rd International Symposium on Molecular Aspects of Chemotherapy (Gdansk, Poland); 5th Molecular Modeling Workshop (Darmstadt, Germany); 11th International Meeting of the Molecular Graphics Society (Chapel Hill, North Carolina); 12th American Peptide Symposium (Cambridge, Massachusetts); 24th Midwest Theoretical Chemistry Conference (DeKalb, Illinois); 32nd American Chemical Society (ACS) National Organic Chemistry Symposium (Minneapolis, Minnesota); 46th Ohio State University International Symposium on Molecular Spectroscopy (Columbus, Ohio); Ab Initio Methods in Quantum Chemistry Symposium Honoring Klaus Reudenberg (Ames, Iowa); ACS Workshop on Molecular Modeling: Methods and Techniques (Athens, Georgia); Computer Prediction of Polymer Properties and Structure (Atlanta, Georgia); Computers in Chemistry (Rochester, New York); Conference on Aspects of Drug Design (Urbana-Champaign, Illinois); Design of Antiviral Agents (Buffalo, New York); Florida School on Applied Molecular Orbital Theory Workshop (Gainesville, Florida); IBM-Polygen Polymer Modeling and Simulation Workshop (Lowell, Massachusetts); Innovative Applications of Computational Chemistry: RISC-Based Parallel Supercomputing (Skokie, Illinois); International Symposium on Computer Simulation of Biomolecular Systems and Mechanisms (Menton, France); Journees de Groupe de Graphism Moleculaire (La Croisic, France); MATH/CHEM/COMP 1991 (Dubrovnik, Croatia); Large Scale Computation for
ings had been back-to-back, a quick traveler could have spent 60 solid days learning about progress in computational chemistry! Although the intensity reached a crescendo in this two-month period, meetings on computational chemistry were held throughout 1991, and 1991 was not atypical of recent years. The field is creating a lot of ripples.

Computational chemistry is "hot" not only because of advances in computer hardware and algorithms, but also because there have been documented (and undocumented) successes, particularly in the pharmaceutical industry. Scientists in adjacent fields view these successes with great expectation. We are living in exciting times!

There are so many developments in the field of computational chemistry that it is difficult to keep track of them. For that reason we established this review series. As in previous volumes, we attempt to treat computational chemistry as broadly and evenhandedly as possible. It should be obvious that not all facets of computational chemistry can be covered in each and every volume. Eventually, however, the existing and future volumes of *Reviews in Computational Chemistry*, when taken in toto, should constitute a useful guide to the field.

We asked the authors of the chapters to begin from ground zero and provide for you a minitutorial on how to implement various computational methods to solve problems. Rather than create a traditional review article, that is, a compilation of data and references that sits in a library, we want you to be able to use this series to learn how to solve problems using computational methods and to be able to locate key references quickly. These chapters are not meant to be surrogate textbooks nor can they replace the original published papers; however, they are meant to be of interest to both experts and novices.
In this volume, several key concepts used by practicing computational chemists are brought into focus. The first chapter by Tamar Schlick is dedicated to the mathematics of optimization. After some mathematical preliminaries, approaches to large-scale optimization are described. Basic decent structure of local methods is highlighted, and then nonderivative, gradient, and Newton methods are explained.

Chapter 2 by Harold Scheraga follows up on the theme of optimization. This chapter on predicting the three-dimensional structures of oligopeptides begins by describing how to construct a model of an oligopeptide chain and then describes methods developed in his laboratory for solving the multiple-minimum problem. This is followed by an extension to larger polypeptides and proteins. The multiple-minima problem, which used to look so knotty, is finally being unraveled.

The third chapter maintains the theme of determining conformations by describing how to generate initial structures of organic and bioorganic molecules and how to model experimental NMR data. Andrew Torda and Wilfred van Gunsteren also discuss refinement methods, force fields, systematic errors and biases, and the quality of predicted structures.

In Chapter 4, David Lewis introduces computer-assisted methods in the evaluation of chemical toxicology. He points out that any substance can be toxic, and thus it is the dose of the substance that determines a toxic response. How, then, does one predict toxicity? Lewis examines QSAR methods, pattern recognition techniques, computer modeling, and knowledge-based systems to answer this question. Ideally, one would like to assess toxicity of a structure before the compound is synthesized. To bring all this into focus, emphasis is placed on the cytochromes P450.

An updated, greatly enlarged compendium of software for molecular modeling appears as the Appendix. Programs that run on personal computers, minicomputers, workstations, mainframes, and supercomputers are listed together with some of their features. Telephone numbers and addresses of the vendors and/or developers are provided. To our knowledge, this is the most complete listing of sources of software for computational chemistry anywhere.

It may surprise some readers, but as editors we have essentially no control over the price of each volume, other than by setting an upper limit on the number of pages. We have worked diligently with VCH Publishers to keep the price of each volume as low as possible. This volume and Volume 4 represent an experiment of having fewer chapters, and hence pages, per volume. The last volume of Reviews in Computational Chemistry (Volume 2) grew to be very thick (527 pages) with a corresponding price. Suggestions have been made to have smaller books that students and others could acquire more readily. Also, by having smaller volumes, readers can pick those containing the topics of most interest to them. Hence, Volumes 3 and 4 are about half the thickness of Volume 2. We want the books in this series to be accessible, affordable, and
useful. Your comments directly to us will be appreciated and will help us design the best format for future volumes.

We express our gratitude to the authors for their fine contributions. We encourage the readers of this review series to recommend other topics by writing to us.

Donald B. Boyd and Kenny B. Lipkowitz
Indianapolis
May 1992
Contents

1. Optimization Methods in Computational Chemistry
   Tamar Schlick

   Introduction 1
   Mathematical Preliminaries 3
     Notation 3
     Problem Statement 3
     Matrix Characteristics 4
     Conditions at Minima 5
     Analysis of Functions 8
   Basic Approaches to Large-Scale Optimization 16
     Size and Space Limitations 16
     Search Techniques 16
     Local and Global Methods 18
   Basic Descent Structure of Local Methods 20
     Descent Directions 21
     Line Search and Trust Region Steps 21
     Convergence Criteria 26
     Convergence Characterization 28
   Nonderivative Methods 29
   Gradient Methods 30
     Steepest Descent 30
     Conjugate Gradient 30
     Preconditioning 32
     Nonlinear Conjugate Gradient 34
   Newton Methods 35
     Overview 35
     Discrete Newton 38
     Quasi-Newton 38
     Truncated Newton 43
   Perspective and Computational Examples 47
     Comparisons 47
     Numerical Example I: Rosenbrock Minimization 51
     Numerical Example II: Deoxycytidine 55

ix
2. Predicting Three-Dimensional Structures of Oligopeptides

Harold A. Scheraga

Introduction 73
Theoretical Foundations 75
Generation of Oligopeptide Chain 76
Residue Geometry 76
End-Group Geometry 76
Constructing a Molecule 77
Ring Closure without Symmetry 79
Ring Closure with Symmetry 80
Early Use of Hard-Sphere Potential 81
More Realistic Potentials 82
Potential Functions 82
Optimization Methods 93
Ancillary Techniques 93
Application to Simple Systems 94
Multiple-Minima Problem 102
Build-up Methods 102
Optimization of Electrostatics (Self-Consistent Electric Field) 106
Monte Carlo plus Minimization 106
Electrostatically Driven Monte Carlo 110
Adaptive Importance Sampling Monte Carlo 110
Increase in Dimensionality 115
Deformation of the Potential Energy Hypersurface 116
Mean-Field Theory 118
Simulated Annealing 121
Extension of Methodology to Large Polypeptides and Proteins 123
Build-up Method 123
Build-up with Limited Constraints 123
Calculations with Constraints 126
Use of Homology 127
Pattern-Recognition Importance Sampling Minimization (PRISM) 128
Outlook for the Future 129
Acknowledgments 129
References 130
3. **Molecular Modeling Using Nuclear Magnetic Resonance Data**  
*Andrew E. Torda and Wilfred F. van Gunsteren*

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Introduction</td>
<td>143</td>
</tr>
<tr>
<td>Scope and Definitions</td>
<td>143</td>
</tr>
<tr>
<td>Historical Perspective</td>
<td>144</td>
</tr>
<tr>
<td>Molecular Representation</td>
<td>145</td>
</tr>
<tr>
<td>Generating Initial Structures</td>
<td>146</td>
</tr>
<tr>
<td>Metric Matrix Method</td>
<td>147</td>
</tr>
<tr>
<td>Variable Target Function Method</td>
<td>149</td>
</tr>
<tr>
<td>Other Methods for Generating Initial Structures</td>
<td>151</td>
</tr>
<tr>
<td>Modeling of Experimental Data</td>
<td>152</td>
</tr>
<tr>
<td>Distance Restraints</td>
<td>152</td>
</tr>
<tr>
<td>Averaging over Discrete Conformations</td>
<td>153</td>
</tr>
<tr>
<td>Time-Averaged Distance Restraints</td>
<td>154</td>
</tr>
<tr>
<td>Direct NOE Refinement</td>
<td>155</td>
</tr>
<tr>
<td>Dihedral Angle Restraints</td>
<td>157</td>
</tr>
<tr>
<td>Refinement, Minimization, and Dynamics</td>
<td>158</td>
</tr>
<tr>
<td>Molecular Dynamics</td>
<td>158</td>
</tr>
<tr>
<td>Other Derivative-Based Dynamics Schemes</td>
<td>159</td>
</tr>
<tr>
<td>Other Non-Derivative-Based Schemes</td>
<td>161</td>
</tr>
<tr>
<td>Force Field</td>
<td>161</td>
</tr>
<tr>
<td>Force Field Parameters and Accuracy</td>
<td>161</td>
</tr>
<tr>
<td>Force Field Modifications</td>
<td>162</td>
</tr>
<tr>
<td>Systematic Errors and Biases</td>
<td>163</td>
</tr>
<tr>
<td>Quality of Structures</td>
<td>164</td>
</tr>
<tr>
<td>Future Directions</td>
<td>165</td>
</tr>
<tr>
<td>Acknowledgment</td>
<td>166</td>
</tr>
<tr>
<td>References</td>
<td>166</td>
</tr>
</tbody>
</table>

4. **Computer-Assisted Methods in the Evaluation of Chemical Toxicity**  
*David F. V. Lewis*

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Introduction</td>
<td>173</td>
</tr>
<tr>
<td>Computer-Based Methods for Toxicity Evaluation</td>
<td>176</td>
</tr>
<tr>
<td>Quantitative Structure-Activity Relationship</td>
<td>176</td>
</tr>
<tr>
<td>Pattern Recognition Techniques</td>
<td>177</td>
</tr>
<tr>
<td>Computer Modeling and Knowledge-Based Systems</td>
<td>179</td>
</tr>
<tr>
<td>The Cytochromes P450</td>
<td>182</td>
</tr>
<tr>
<td>COMPACT</td>
<td>190</td>
</tr>
<tr>
<td>Molecular Orbital Calculations and QSARs in Toxicity</td>
<td>200</td>
</tr>
<tr>
<td>Critical Assessment of Predictive Methods</td>
<td>203</td>
</tr>
<tr>
<td>Conclusions and Future Prospects</td>
<td>208</td>
</tr>
</tbody>
</table>
Information Sources for Software for Computer-Aided Prediction of Toxicity Methods 210
Acknowledgments 211
References 211

Appendix: Compendium of Software for Molecular Modeling
Donald B. Boyd

Introduction 223
References 225
Software for Personal Computers 226
   General Purpose Molecular Modeling 226
   Quantum Chemistry Calculations 229
   Databases of Molecular Structures 230
   Molecular Graphics and Other Applications 230
Software for Minicomputers, Superminicomputers, Workstations, and Supercomputers
   General-Purpose Molecular Modeling 233
   Quantum Chemistry Calculations 239
   Databases of Molecular Structures 243
   Molecular Graphics and Other Applications 246

Author Index 249
Subject Index 261
Contributors

Donald B. Boyd, Lilly Research Laboratories, Eli Lilly and Company, Lilly Corporate Center, Indianapolis, Indiana 46285, U.S.A. (Electronic mail: boyd.donald.b@lilly.com)

David F. V. Lewis, Molecular Toxicology Research Group, Division of Toxicology, School of Biological Sciences, University of Surrey, Guildford, Surrey, GU2 5XH, United Kingdom

Harold A. Scheraga, Baker Laboratory of Chemistry, Cornell University, Ithaca, New York 14853-1301, U.S.A. (Electronic mail: mwv@cornellc.bitnet)

Tamar Schlick, Courant Institute of Mathematical Sciences and Chemistry Department, New York University, 251 Mercer Street, New York, New York, 10012, U.S.A. (Electronic mail: schlick@acfclu.nyu.edu)

Andrew E. Torda, Physical Chemistry, ETH Zentrum, CH 8092, Zürich, Switzerland (Electronic mail: torda@igc.ethz.ch)

Wilfred F. van Gunsteren, Physical Chemistry, ETH Zentrum, CH 8092, Zürich, Switzerland (Electronic mail: wfvgm@igc.ethz.ch)
Contributors to Previous Volumes

Volume I

David Feller and Ernest R. Davidson, Basis Sets for *Ab Initio* Molecular Orbital Calculations and Intermolecular Interactions.

James J. P. Stewart, Semiempirical Molecular Orbital Methods.

Clifford E. Dykstra, Joseph D. Augspurger, Bernard Kirtman, and David J. Malik, Properties of Molecules by Direct Calculation.


Peter C. Jurs, Chemometrics and Multivariate Analysis in Analytical Chemistry.

Yvonne C. Martin, Mark G. Bures, and Peter Willett, Searching Databases of Three-Dimensional Structures.

Paul G. Mezey, Molecular Surfaces.


Donald B. Boyd, Aspects of Molecular Modeling.

Donald B. Boyd, Successes of Computer-Assisted Molecular Design.

Ernest R. Davidson, Perspectives on *Ab Initio* Calculations.

*Current address: University of Washington, Seattle.
Volume II

Andrew R. Leach,† A Survey of Methods for Searching the Conformational Space of Small and Medium-Sized Molecules.


Steve Scheiner, Calculating the Properties of Hydrogen Bonds by Ab Initio Methods.

Donald E. Williams, Net Atomic Charge and Multipole Models for the Ab Initio Molecular Electric Potential.

Peter Politzer and Jane S. Murray, Molecular Electrostatic Potentials and Chemical Reactivity.

Michael C. Zerner, Semiempirical Molecular Orbital Methods.


Donald B. Boyd, The Computational Chemistry Literature.

†Current address: University of Southampton, U.K.
CHAPTER 1

Optimization Methods in Computational Chemistry

Tamar Schlick

Courant Institute of Mathematical Sciences and Chemistry Department, New York University, 251 Mercer Street, New York, New York 10012

INTRODUCTION

In his 1970 *Numerical Methods*, the witty numerical analyst Forman Acton wrote:

It is with a sense of reluctance that your author introduces this topic, for minimum-seeking methods are often used when a modicum of thought would disclose more appropriate techniques. They are the first refuge of the computational scoundrel, and one feels at times that the world would be a better place if they were quietly abandoned. But even if these techniques are frequently misused, it is equally true that there are problems for which no alternative solution method is known—and so we shall discuss them. A better title for this chapter might be "How to Find Minima—If You Must!"  

Although there may be some truth in this view, there is little doubt today that multivariate minimization algorithms are fundamental research tools for scientists in numerous disciplines. A common problem arises when a complex physical system is described by a collection of particles, or combinations of states,
a multidimensional phase space. An energy or cost function is associated with each different configuration, and the challenge is to find sets of points that minimize (or maximize) the objective function. Such applications arise frequently in molecular modeling, rational drug design, mathematical biology models, neural networks, combinatorial problems, financial investment planning, architectural design, electronics, meteorology, and computational geometry. I wonder: Would Acton still call us all scoundrels?

The purpose of this chapter is to provide computational chemists a brief background into unconstrained nonlinear optimization methods that seek local minima. Emphasis is placed on methods that are most powerful today for large-scale problems (hundreds to thousands of independent variables) and suitable for potential energy minimization. Only a general practical taste of this very rich and theoretically interesting field is attempted here. For comprehensive treatments of optimization methods, interested readers are referred to a large selection of excellent books on a wide spectrum of levels, from introductory\textsuperscript{1,2} to comprehensive\textsuperscript{3–7} to specialized-topic volumes.\textsuperscript{8–12} The numerical linear algebra and multivariate calculus background can be found in most of the introductory optimization books as well as in standard books on numerical methods and matrix computations.\textsuperscript{13,14} Many general concepts mentioned throughout this chapter without explicit citations are discussed in these volumes; however, an attempt is made to make this review as self-contained as possible.

Another introductory note may provide further incentive for novice optimizers to read on. Despite extensive developments in optimization methods in the last decade, large-scale nonlinear optimization still remains an art that requires considerable computing experience, algorithm familiarity, and intuition. In general, “black box” minimization implementations, even those using state-of-the-art algorithms, are only partially successful.

“I have been burnt by too many black boxes,” declares Acton in his 1990 preface.\textsuperscript{1} There are at least two reasons for our shared sentiment. First, many commercial software vendors, such as NAG and Harwell, have developed general-purpose programs that are, for the most part, easy to use but not always up-to-date with the latest minimization developments. Thus, successful new minimization approaches are often out of reach for nonspecialist mathematicians, let alone scientists in related application fields. Second, application tailoring of algorithms is often very difficult with ready-made software. Such modifications may be crucial in many applications.

For example, a common thread in applications in meteorology, chemistry, or mathematical biology is a natural separability of the objective functions into components of differing complexity (e.g., local and nonlocal interactions). This composition may not only change the relative attractiveness or suitability among different optimization methods, but also lead to very powerful methods for the application at hand when this information is incorporated appropriately.
Mathematical Preliminaries

Such problem tailoring requires some familiarity with the algorithmic modules. It also demands knowledge of the theoretical and practical strengths and weaknesses of the different minimization methods. With rapidly growing improvements in high-performance super and massively parallel machines, application-tailored software may be even more important in combination with parallel architectures whose design is motivated by specific applications.

MATHEMATICAL PRELIMINARIES

Notation

We generally denote scalars by lowercase Greek letters (e.g., $\beta$), column vectors by boldface lowercase Roman letters (e.g., $x$), and matrices by capital italic Roman letters (e.g., $H$). A superscript $^T$ denotes a vector or matrix transpose. Thus $x^T$ is a row vector, $x^T y$ is an inner product, and $A^T$ is the transpose of the matrix $A$. Unless stated otherwise, all vectors belong to $\mathbb{R}^n$, the $n$-dimensional vector space. Components of a vector are typically written as italic letters with subscripts (e.g., $x_1, x_2, \ldots, x_n$). The standard basis vectors in $\mathbb{R}^n$ are the $n$ vectors $\{e_1, e_2, \ldots, e_n\}$, where $e_j$ has the entry 1 in the $j$th component and 0 in all others. Often, the associated vector norm is the standard Euclidean norm, $\|\cdot\|_2$, defined as

$$
\|x\| = \left( \sum_{i=1}^n |x_i|^2 \right)^{1/2}.
$$

We say that a term is of order $n$ and write $O(n)$ to mean that it is proportional to $n$.

Problem Statement

We are interested in solving the optimization problem without constraints:

$$
\text{minimize } f(x), \quad x \in \mathbb{R}^n
$$

where $f$ is a real-valued function of $n$ variables. We will assume that $f$ is sufficiently smooth, that is, possesses continuous derivatives at least up to second order.

The gradient of $f$ at $x$ is defined to be the first derivative vector $\nabla f(x)$, or $g(x)$, whose $n$ components are given by $g_i(x) = \partial E(x)/\partial x_i$. The Hessian at $x$, $H(x)$, is defined to be the $n \times n$ matrix of second partial derivatives with components $H_{ij}(x) = [\partial^2 f(x)/\partial x_i \partial x_j]$. As $\partial^2 f(x)/\partial x_i \partial x_j = \partial^2 f(x)/\partial x_j \partial x_i$, the Hessian matrix is symmetric: $H_{ij}(x) = H_{ji}(x)$. 
For example, for the following function of two variables,
\[ f(x) = e^{x_1}(4x_1^2 + 4x_1x_2 + 2x_2^2), \]
the gradient vector is
\[ g(x) = \begin{bmatrix} 2e^{x_1}(2x_1^2 + 2x_1x_2 + x_2^2 + 4x_1 + 2x_2) \\ 4e^{x_1}(x_1 + x_2) \end{bmatrix} \]
and the Hessian is the matrix
\[ H(x) = \begin{bmatrix} 2e^{x_1}(2x_1^2 + 2x_1x_2 + x_2^2 + 8x_1 + 4x_2 + 4) & 4e^{x_1}(1 + x_1 + x_2) \\ 4e^{x_1}(1 + x_1 + x_2) & 4e^{x_1} \end{bmatrix}. \]

Just as in the one-dimensional case the derivative defines the slope of the tangent line to the curve \( f(x) \), the gradient vector at \( x \) represents the normal to the tangent hyperplane at the point \( x \). The term \textit{hyperplane} is an extension to \( n > 3 \) of a plane in three dimensions. All vectors \( x, y \) satisfying \( x^Ty = \gamma \) for some constant \( \gamma \) lie in a hyperplane in higher dimensions. Thus, all vectors \( y \) satisfying \( g(x)^Ty = \gamma \) where \( \gamma = g(x)^Tx \) lie in the tangent hyperplane at \( x \).

A point \( x^* \in \mathbb{R}^n \) is said to be a \textit{strict local minimum} of \( f \) if there exists a number \( \eta > 0 \) such that \( f(x^*) < f(x) \) for all \( x \neq x^* \) within a given distance \( \eta \) from \( x^* \) (i.e., \( \|x - x^*\| < \eta \)). For a \textit{weak minimum}, only \( f(x^*) \leq f(x) \) holds. The point \( x^* \) is a \textit{global minimum} of \( f \) if \( f(x^*) < f(x) \) for all \( x \neq x^* \), \( x \in \mathbb{R}^n \). Figure 1 illustrates these possibilities for a one-dimensional function.

**Matrix Characteristics**

There are several general characteristics of a matrix that are particularly useful for analysis of minimization algorithms. Density of a matrix is a measurement given by the ratio of the nonzero to zero matrix components. A matrix is said to be dense when this ratio is large and sparse when it is small. A sparse matrix may be structured (e.g., block diagonal, band) or unstructured (Figure 2).

A symmetric matrix \( A \) is said to be positive-definite if the quadratic form \( u^TAu > 0 \) for all nonzero vectors \( u \). Similarly, the symmetric matrix \( A \) is positive-semidefinite if \( u^TAu \geq 0 \) for all nonzero vectors \( u \). Positive-definite matrices have strictly positive eigenvalues. We classify \( A \) as negative-definite if \( u^TAu < 0 \) for all nonzero vectors \( u \). \( A \) is indefinite if \( u^TAu \) is positive for some \( u \) and negative for others.

For example, the function \( f(x) \) defined in Eq. [3] can be written as the product of \( e^{x_1} \) and a generalized quadratic function: \( f(x_1,x_2) = e^{x_1}[x^TAx] \). The matrix \( A = \begin{bmatrix} 4 & 2 \\ 2 & 3 \end{bmatrix} \) is positive-definite.
The Hessian matrix is a generalization in $\mathbb{R}^n$ of the concept of curvature of a function. The positive-definiteness of the Hessian is a generalized notion of positive curvature. Thus, the properties of $H$ are very important in formulating minimum-seeking algorithms.

Higher-order derivatives are rarely used in minimization methods; however, some recent approaches, termed tensor, have attempted to approximate higher-order information cheaply.$^{17}$

**Conditions at Minima**

What are the sufficient conditions that must hold at a solution point of problem [2]? These conditions are simply extensions to $\mathbb{R}^n$ of the well-known first and second derivative conditions for univariate functions. Let us assume that $f(x)$ is a smooth function with continuous first, second, and third derivatives defined for all $x$. We further suppose that the following conditions hold for the gradient and Hessian of $f$ at $x^* \in \mathbb{R}^n$:

(i) $g(x^*) = 0$,  
(ii) $H(x^*)$ is positive-definite.

Then it can be easily shown that $x^*$ is a strict local minimum of $f$.

The Taylor expansion of $f(x)$ about $x^*$ along a perturbation vector $p = (\Delta x_1, \Delta x_2, \ldots, \Delta x_n)$ produces
Figure 2 Sample matrix patterns for (a) block diagonal and (b–e) sparse unstructured. Pattern (b) corresponds to the Hessian approximation (preconditioner) for a potential energy model from the local energy terms (bond length, bond angle, and dihedral angle terms), and (c) is a reordered matrix pattern that reduces fill-in during the factorization. Pattern (d) comes from a molecular dynamics simulation of supercoiled DNA and describes pairs of points along a ribbonlike model of the duplex that come in close contact during the dynamics trajectory; pattern (e) is the associated reordered structure that reduces fill-in.

\[ f(x^* + p) = f(x^*) + \sum_{i=1}^{n} (\Delta x_i) \frac{\partial f(x^*)}{\partial x_i} + \frac{1}{2} \sum_{i=1}^{n} (\Delta x_i^2) \frac{\partial^2 f(x^*)}{\partial x_i^2} \]

\[ + \sum_{i=1}^{n} \sum_{j>i} (\Delta x_i)(\Delta x_j) \frac{\partial^2 f(x^*)}{\partial x_i \partial x_j} + \text{higher order terms.} \]  

[6]

More compactly, this expansion can be written as

\[ f(x^* + p) = f(x^*) + g(x^*)^T p + \frac{1}{2} p^T H(x^*) p + O(\|p\|^3). \]  

[7]
As condition (i) holds, the gradient term vanishes, and condition (ii) implies that there exists some \( \alpha > 0 \) for which \( p^T H(x^*) p > \alpha \|p\|^2 \). It follows that we can write

\[
f(x^* + p) - f(x^*) \geq \frac{\alpha}{2} \|p\|^2 + O(\|p\|^3).
\]

For arbitrary small perturbation vectors \( \|p\| \), the first term on the right dominates the second; consequently, the right-hand side is positive. By definition, \( x^* \) is a strict local minimum of \( f \). \( \Box \)

A point \( x^* \) is called a stationary point of \( f \) if \( g(x^*) = 0 \) but \( H(x^*) \) is not necessarily positive-definite. Thus, local and global minima are stationary points, but there are more general stationary points, such as saddle points, which are neither local nor global minima. Special techniques are needed for detection of saddle points, which are often related to structural transitions in molecular applications.
Figure 3 illustrates the different types of stationary points for one-dimensional functions. The minimum and maximum at the origin are shown, respectively, for a convex and concave quadratic function. Note that the traces in the $x,z$ and $y,z$ planes are parabolas, opening upward (above origin) in the case of the minimum and opening downward (below origin) for the maximum. Elliptical cross sections can be noted in the $x,y$ planes. The saddle point at the origin exhibits parabolic traces in the $x,z$ and $y,z$ planes, opening upward and downward, respectively. The trace in the $x,y$ plane is a pair of intersecting lines.

**Analysis of Functions**

The positive-definiteness of $H$ is a useful concept in analysis of general functions. Smooth functions can be approximated by quadratic models within a sufficiently small neighborhood of a given point. The local behavior of $f$ can then be analyzed in terms of the properties of $H$. 
To see this, consider the Taylor expansion of a quadratic function $q$ about a stationary point $x^*$:

$$q(x^* + p) = q(x^*) + \frac{1}{2}p^T H(x^*) p.$$  \[9\]

The symmetry of $H$ implies that the $n$ distinct eigenvectors $\{v_i\}$ associated with the eigenvalues $\{\lambda_i\}$ are orthonormal (i.e., $v_i^T v_j = 0$ for $i \neq j$, and $v_i^T v_i = 1$ for all $i$). We can then express $p$ as the linear combination of eigenvectors

$$p = \sum_{i=1}^{n} \alpha_i v_i.$$  \[10\]

for a scalar set $\{\alpha_i\}$. As $v_i^T H v_i = \lambda_i$ for each $i$, the difference in function value caused by a movement along $p$ is given by
Clearly, the signs of the eigenvalues influence the change in $q$ along particular directions. For example, when $p = \alpha v_i$ for $\alpha > 0$ and some $i$, $q$ will be strictly increasing if $\lambda_i > 0$ and strictly decreasing if $\lambda_i < 0$ as $\alpha$ increases, respectively. If $\lambda_i = 0$, $q$ is a constant along directions parallel to $v_i$ (because $Hv_i = 0$) and reduces to a linear function along this direction (the quadratic term in Eq. [9] vanishes). When all the eigenvalues are positive, $x^*$ is a unique global minimum of $q$. If $H$ is positive-semidefinite (i.e., has nonnegative eigenvalues), $x^*$ is a weak minimum. If $H$ is indefinite and nonsingular (i.e., no eigenvalues are zero), the critical point $x^*$ is a saddle point.

For example, the two-dimensional function defined in Eq. [3] has a local minimum at (0,0) and a saddle point at (-2,2). Both points are stationary. At
Figure 3  Types of stationary points.