Appendix A
Solving Some Auxiliary Optimization Problems

A.1 Newton's Method for Univariate Minimization

Let us show that Newton's Method is very efficient in finding the maximal root of increasing convex univariate functions. Consider a univariate function f such that

$$f(\tau_*) = 0, \quad f(\tau) > 0, \text{ for } \tau > \tau_*, \tag{A.1.1}$$

and it is convex for $\tau \geq \tau_*$. Let us choose $\tau_0 > \tau_*$. Consider the following Newton process:

$$\tau_{k+1} = \tau_k - \frac{f(\tau_k)}{g_k}, \tag{A.1.2}$$

where $g_k \in \partial f(\tau_k)$. Thus, we do not assume f to be differentiable for $\tau \geq \tau_*$.

Theorem A.1.1 *Method (A.1.2) is well defined. For any $k \geq 0$ we have*

$$f(\tau_{k+1})g_{k+1} \leq \frac{1}{4} f(\tau_k)g_k. \tag{A.1.3}$$

Thus, $f(x_k) \leq \left(\frac{1}{2}\right)^k g_0(\tau_0 - \tau_*)$.

Proof Let $f_k = f(\tau_k)$. Let us assume that $f_k > 0$ for all $k \geq 0$. Since f is convex for $\tau \geq \tau_*$, $0 = f(\tau_*) \geq f_k + g_k(\tau_* - \tau_k)$. Thus,

$$g_k(\tau_k - \tau_*) \geq f_k > 0. \tag{A.1.4}$$

This means that $g_k > 0$ and $\tau_{k+1} \in [\tau_*, \tau_k)$. In particular, we conclude that

$$\tau_k - \tau_* \leq \tau_0 - \tau_*. \tag{A.1.5}$$

Y. Nesterov, *Lectures on Convex Optimization*, Springer Optimization and Its Applications 137, https://doi.org/10.1007/978-3-319-91578-4

Further, for any $k \geq 0$ we have:

$$f_k \geq f_{k+1} + g_{k+1}(\tau_k - \tau_{k+1}) \overset{(A.1.2)}{=} f_{k+1} + \frac{f_k g_{k+1}}{g_k}.$$

Thus, $1 \geq \frac{f_{k+1}}{f_k} + \frac{g_{k+1}}{g_k} \geq 2\sqrt{\frac{f_{k+1}g_{k+1}}{f_k g_k}}$, and this is (A.1.3). Finally, since f is convex for $\tau \geq \tau_*$, we have

$$g_0 \overset{(A.1.4)}{\geq} \sqrt{\frac{f_0 g_0}{\tau_0 - \tau_*}} \overset{(A.1.3)}{\geq} 2^k \sqrt{\frac{f_k g_k}{\tau_0 - \tau_*}} \overset{(A.1.4)}{\geq} 2^k \sqrt{\frac{f_k^2}{(\tau_0 - \tau_*)(\tau_k - \tau_*)}}$$

$$\overset{(A.1.5)}{\geq} 2^k \frac{f_k}{\tau_0 - \tau_*}. \qquad \square$$

Thus, we have seen that method (A.1.2) has linear rate of convergence, which does not depend on the particular properties of the function f. Let us show that in a non-degenerate situation this method has local quadratic convergence.

Theorem A.1.2 *Let a convex function f be twice differentiable. Assume that it satisfies the conditions (A.1.1) and its second derivative increases for $\tau \geq \tau_*$. Then for any $k \geq 0$ we have*

$$f(\tau_{k+1}) \leq \frac{f''(\tau_k)}{2(f'(\tau_k))^2} \cdot f^2(\tau_k). \tag{A.1.6}$$

If the root τ_* is non-degenerate:

$$f'(\tau_*) > 0, \tag{A.1.7}$$

then $f(\tau_{k+1}) \leq \frac{f''(\tau_0)}{2(f'(\tau_*))^2} \cdot f^2(\tau_k)$.

Proof In view of conditions of the theorem, $f''(\tau) \leq f''(\tau_k)$ for all $\tau \in [\tau_{k+1}, \tau_k]$. Therefore,

$$\begin{aligned} f(\tau_{k+1}) &\leq f(\tau_k) + f'(\tau_k)(\tau_{k+1} - \tau_k) + \tfrac{1}{2} f''(\tau_k)(\tau_{k+1} - \tau_k)^2 \\ &\overset{(A.1.2)}{=} \tfrac{1}{2} f''(\tau_k) \frac{f^2(\tau_k)}{(f'(\tau_k))^2}. \end{aligned}$$

To prove the last statement, it remains to note that $f''(\tau_k) \leq f''(\tau_0)$ and $f'(\tau_k) \geq f'(\tau_*)$. $\square$

A.2 Barrier Projection onto a Simplex

In the case $K = \mathbb{R}^n_+$, we can take

$$F(x) \;=\; -\sum_{i=1}^{n} \ln x^{(i)}, \qquad \nu = n.$$

Consider $\hat{P} = \{x \in \mathbb{R}^n_+ : \; \langle \bar{e}_n, x\rangle = 1\}$. Then, at each iteration of method (7.3.14) we need to solve the following problem:

$$\phi^* \stackrel{\mathrm{def}}{=} \max_x \left\{ \langle s, x\rangle + \sum_{i=1}^{n} \ln x^{(i)} : \; \sum_{i=1}^{n} x^{(i)} = 1 \right\}. \tag{A.2.1}$$

Let us show that its complexity does not depend on the size of particular data (that is, the coefficients of the vector $s \in \mathbb{R}^n$).

Consider the following Lagrangian:

$$\mathscr{L}(x,\lambda) = \langle s, x\rangle + \sum_{i=1}^{n} \ln x^{(i)} + \lambda \cdot \left[1 - \sum_{i=1}^{n} x^{(i)} \right], \qquad x \in \mathbb{R}^n, \;\; \lambda \in \mathbb{R}.$$

The dual function

$$\phi(\lambda) = \max_x \left\{ \mathscr{L}(x,\lambda) : \; \sum_{i=1}^{n} x^{(i)} = 1 \right\} \stackrel{\mathrm{def}}{=} \mathscr{L}(x(\lambda),\lambda)$$

is defined by the vector $x(\lambda) : \; x^{(i)}(\lambda) = \frac{1}{\lambda - s^{(i)}}$, $i = 1, \dots, n$. Thus,

$$\begin{aligned} \phi(\lambda) &= -n + \lambda - \sum_{i=1}^{n} \ln\left(\lambda - s^{(i)}\right), \\ \phi_* &= \min_\lambda \left\{ \phi(\lambda) : \; \lambda > \max_{1 \le i \le n} s^{(i)} \right\}. \end{aligned} \tag{A.2.2}$$

Note that $\phi(\cdot)$ is a standard self-concordant function. Therefore we can apply to its minimization the intermediate Newton's Method (5.2.1), Item C), which converges quadratically starting from any λ from the region

$$\mathscr{Q}(s) \;=\; \{\lambda : \; 4(\phi'(\lambda))^2 \le \phi''(\lambda)\}$$

(see Theorem 5.2.2). Let us show that the complexity of finding a starting point from this set does not depend on the initial data.

Consider the function $\psi(\lambda) \;=\; -\phi'(\lambda) \;=\; \sum_{i=1}^{n} \frac{1}{\lambda - s^{(i)}} - 1$. Clearly, the problem (A.2.2) is equivalent to finding the largest root λ_* of the equation

$$\psi(\lambda) \;=\; 0. \tag{A.2.3}$$

Let $\lambda_0 = 1 + \max\limits_{1 \le i \le n} s^{(i)}$. Then $\psi(\lambda_0) \ge 0$ and therefore $\lambda_0 \le \lambda_*$. Consider the following process:

$$\lambda_{k+1} = \lambda_k - \frac{\psi(\lambda_k)}{\psi'(\lambda_k)}, \quad k \ge 0. \tag{A.2.4}$$

This is a standard Newton's method for solving the Eq. (A.2.3), which can be also interpreted as a Newton's method for the minimization problem (A.2.2).

Lemma A.2.1 *For any* $k \ge 0$ *we have* $(\phi'(\lambda_k))^2 \le n^7 \cdot \left(\frac{1}{16}\right)^k \phi''(\lambda_k)$.

Proof Note that function ψ is decreasing and strictly convex. Therefore, for any $k \ge 0$ we have

$$\lambda_k < \lambda_{k+1} < \lambda_*, \quad \psi'(\lambda_k) < 0, \quad \psi(\lambda_k) > 0.$$

Since $\psi(\lambda_k) \ge \psi(\lambda_{k+1}) + \psi'(\lambda_{k+1})(\lambda_k - \lambda_{k+1}) = \psi(\lambda_{k+1}) + \frac{\psi'(\lambda_{k+1})}{\psi'(\lambda_k)}\psi(\lambda_k)$, we obtain[1]

$$1 \ge \frac{\psi(\lambda_{k+1})}{\psi(\lambda_k)} + \frac{\psi'(\lambda_{k+1})}{\psi'(\lambda_k)} \ge 2\sqrt{\frac{\psi(\lambda_{k+1})\psi'(\lambda_{k+1})}{\psi(\lambda_k)\psi'(\lambda_k)}}.$$

Thus, for any $k \ge 0$ we get

$$\phi''(\lambda_k) \cdot |\phi'(\lambda_k)| \le \left(\frac{1}{4}\right)^k \phi''(\lambda_0) \cdot |\phi'(\lambda_0)|. \tag{A.2.5}$$

Further, in view of the choice of λ_0 we have

$$|\phi'(\lambda_0)| = \psi(\lambda_0) = \sum_{i=1}^{n} \frac{1}{\lambda_0 - s^{(i)}} - 1 < n - 1,$$

$$\phi''(\lambda_0) = \sum_{i=1}^{n} \frac{1}{(\lambda_0 - s^{(i)})^2} \le n.$$

Finally, since $0 \le \psi(\lambda_k) = \sum\limits_{i=1}^{n} \frac{1}{\lambda_k - s^{(i)}} - 1$, we conclude that

$$\phi''(\lambda_k) = \sum_{i=1}^{n} \frac{1}{(\lambda_k - s^{(i)})^2} \ge \frac{1}{n}.$$

[1]We use the same arguments as in the proof of Theorem A.1.1, but for a decreasing univariate function.

Using these bounds in (A.2.5), we obtain

$$\frac{1}{\phi''(\lambda_k)}(\phi'(\lambda_k))^2 \leq \left(\frac{1}{16}\right)^k \frac{(\phi''(\lambda_0))^2(\phi'(\lambda_0))^2}{(\phi''(\lambda_k))^3} \leq \left(\frac{1}{16}\right)^k \cdot n^7. \qquad \square$$

Comparing the statement of Lemma A.2.1 with the definition of $\mathscr{Q}(s)$, we conclude that the process (A.2.4) arrives at the region of quadratic convergence at most after

$$\left\lceil \tfrac{1}{4}(2 + 7\log_2 n) \right\rceil \tag{A.2.6}$$

iterations. Each such iteration takes $O(n)$ arithmetic operations.

A similar technique can be used for finding the barrier projection in the cone of positive-semidefinite matrices:

$$\max_X \{\langle S, X\rangle + \ln\det X : \langle I_n, X\rangle = 1\}.$$

The most straightforward strategy consists in finding an eigenvalue decomposition of the matrix S and solving the problem (A.2.1) with s being the spectrum of the matrix. In a more efficient strategy, we transform S into tri-diagonal form by an orthogonal transformation, compute its maximal eigenvalue and apply the Newton's method to the corresponding dual function.

Bibliographical Comments

In the past few decades, numerical methods for Convex Optimization have become widely studied in the monographic literature. The reader interested in engineering applications can benefit from the introductory exposition by Polyak [55], excellent course by Boyd and Vandenberghe [6], and lecture notes by Ben-Tal and Nemirovski [5]. Mathematical aspects are described in detail in the older lectures by A. Nemirovski (see [33] for the Internet version) and in the original versions of the theory for Interior-Point Methods by Renegar [57], Roos et al. [59], and Ye [63]. Recent theoretical highlights can be found in the monographs by Beck [3] and Bubeck [7]. In our book, we have tried to be more balanced, combining the comprehensive mathematical theory with many examples of practical applications, sometimes supported by numerical experiments.

Chapter 1: Nonlinear Optimization

Section 1.1 The complexity theory for black-box optimization schemes was developed in [34], where the reader can find different examples of resisting oracles and lower complexity bounds similar to that of Theorem 1.1.2.

Sections 1.2 and 1.3 There exist several classical monographs [11, 12, 30, 53] treating different aspects of Nonlinear Optimization. For understanding Sequential Unconstrained Minimization, the best source is still [14]. Some facts in Sect. 1.3, related to conditions for zero duality gap, are probably new.

Y. Nesterov, *Lectures on Convex Optimization*, Springer Optimization and Its Applications 137, https://doi.org/10.1007/978-3-319-91578-4

Chapter 2: Smooth Convex Optimization

Section 2.1 The original lower complexity bounds for smooth convex and strongly convex functions can be found in [34]. The proof used in this section was first published in [39].

Section 2.2 Gradient mapping was introduced in [34]. The first optimal method for smooth and strongly convex functions was proposed in [35]. The constrained variant of this scheme is taken from [37]. However, the framework of estimating sequences was suggested for the first time in [39]. A discussion of different approaches for generating points with small norm of the gradient can be found in [48].

Section 2.3 Optimal methods for discrete minimax problems were developed in [37]. The approach of Sect. 2.3.5 was first described in [39].

Chapter 3: Nonsmooth Convex Optimization

Section 3.1 A comprehensive treatment of different topics of Convex Analysis can be found in [24]. However, the classical monograph [58] is still very useful.

Section 3.2 Lower complexity bounds for nonsmooth minimization problems can be found in [34]. The framework of Sect. 3.2.2 was suggested in [36]. For detailed bibliographical comments on the early history of Nonsmooth Minimization see [55, 56].

Section 3.3 The example of a difficult function for Kelley's method is taken from [34]. The presentation of the Level Method in this section is close to [28].

Chapter 4: Second-Order Methods

Section 4.1 Starting from the seminal papers of Bennet [4] and Kantorovich [26], Newton's Method became an important tool for numerous applied problems. In the last 50 years, the number of different suggestions for improving the scheme is extremely large (see, for example, [11, 12, 15, 21, 29, 31]). The reader can consult an exhaustive bibliography in [11].

Most probably, the natural idea of using cubic regularization to improve the stability of the Newton scheme was first analyzed in [22]. However, the author was very sceptical about the complexity of solving the auxiliary minimization problem in the case of nonconvex quadratic approximation (and indeed, it can have an exponential number of local minima). As a result, this paper was never published. Twenty five years later, in an independent paper [52] this idea was checked again, and it was shown that this problem is solvable by standard techniques of Linear Algebra. The authors also developed global worst-case complexity bounds for different problem classes. This paper forms the basis of Sect. 4.1. The interested reader can also consult the complementary approach [8, 9], where cubic regularization is coupled with a line search along the gradient direction. However, note that this feature, though improving somewhat the numerical stability, forces the algorithm to stop at saddle points. A historical exposition of the development in this field with recent results, including lower complexity bounds for gradient norm minimization, can be found in [10].

Section 4.2 This section is based on the paper [45].

Section 4.3 This section is based on very recent and partially unpublished results. The first lower complexity bounds for second-order methods were obtained in [2]. At the same time, one of the second-order schemes in [32] achieves the rate of convergence $\tilde{O}\left(\frac{1}{k^{7/2}}\right)$, which is optimal. However, each iteration of this method needs an expensive search procedure based on additional calls of oracle. So, its practical efficiency is questionable.

In our presentation, we use a simpler derivation of the lower complexity bounds and a simpler conceptual version of the "optimal" second-order scheme, based on iteration of the Cubic Newton Method.

Section 4.4 Methods for solving systems of nonlinear equations have attracted a lot of attention (see [11, 12, 53, 54]). However, we have not been able to find any global worst-case efficiency estimates for them in the literature. Our presentation follows the paper [43].

Chapter 5: Polynomial-Time Interior-Point Methods

This chapter contains an adaptation of the main concepts from [51]. We added several useful inequalities and a slightly simplified presentation of the path-following scheme. We refer the reader to [5] for numerous applications of interior-point methods, and to [57, 59, 62] and [63] for a detailed treatment of different theoretical aspects.

Section 5.1 In this section, we introduce the definition of a self-concordant function and study its properties. As compared with Section 4.1 in [39], we add Fenchel duality and the Implicit Function Theorem. The main novelty is an explicit treatment of the constant of self-concordance. However, most of the material can be found in [51].

Section 5.2 In this new section, we analyze different methods for minimizing self-concordant functions. We propose a new step-size rule for the Newton scheme (*intermediate step*), which gives better constants for the path-following approach. Complexity estimates for a path-following scheme, as applied to a self-concordant function, were obtained only recently [13].

Section 5.3 In this section we study the properties of a self-concordant barrier and give the complexity analysis for the path-following method. This is an adaptation of Section 4.2 in [39].

Section 5.4 In this section, we give examples of self-concordant barriers and related applications. This is an extension of Section 4.3 in [39] by the results of [49].

Chapter 6: The Primal-Dual Model of an Objective Function

This is the first attempt at presenting in the monographic literature the fast primal-dual gradient methods based on an explicit minimax model of the objective function. In the first three sections we present different aspects of the smoothing technique, following the papers [40, 41], and [42]. It seems that the Fast Gradient Method in the form of the Method of Similar Triangles (6.1.19) was published for the first time only recently (see [20]).

The last Sect. 6.4 is devoted to the new analysis of the old Conditional Gradient Method (or, the *Frank–Wolfe algorithm* [16, 18, 19, 23, 25]). Our presentation follows the paper [50], which is close in spirit to [17].

Chapter 7: Optimization in Relative Scale

The presentation in this new chapter is based on the papers [44, 46], and [47]. Some examples of application were analyzed in [5], however, from the viewpoint of the applicability of Interior-Point Methods. Algorithms for computing the rounding ellipsoids are studied in [1, 27, 61], and in the recent book [60]. Constant quality of semidefinite relaxation for Boolean quadratic maximization with general matrix was proved in [38]. The material of Sect. 7.4 is new.

References

1. K.M. Anstreicher, Ellipsoidal approximations of convex sets based on the volumetric barrier. CORE Discussion Paper 9745, 1997
2. Y. Arjevani, O. Shamir, R. Shiff, Oracle complexity of second-order methods for smooth convex optimization. arXiv:1705.07260v2 (2017)
3. A. Beck, *First-Order Methods in Optimization* (SIAM, Philadelphia, 2017)
4. A.A. Bennet, Newton's method in general analysis. Proc. Natl. Acad. Sci. U. S. A. **2**(10), 592–598 (1916)
5. A. Ben-Tal, A. Nemirovskii, *Lectures on Modern Convex Optimization: Analysis, Algorithms, and Engineering Applications* (SIAM, Philadelphia, 2001)
6. S. Boyd, L. Vandenberghe, *Convex Optimization* (Cambridge University Press, Cambridge, 2004)
7. S. Bubeck, *Convex Optimization: Algorithms and Complexity* (Now Publishers, LP Breda, 2015). arXiv:1405.4980
8. C. Cartis, N.I.M. Gould, P.L. Toint, Adaptive cubic regularisation methods for unconstrained optimization. Part I: motivation, convergence and numerical results. Math. Program. **127**(2), 245–295 (2011)
9. C. Cartis, N.I.M. Gould, P.L. Toint, Adaptive cubic regularisation methods for unconstrained optimization. Part II: worst-case function- and derivative-evaluation complexity. Math. Program. **130**(2), 295–319 (2011)
10. C. Cartis, N.I.M. Gould, P.L. Toint, How much patience do you have? a worst-case perspective on smooth nonconvex optimization. Optima **88**, 1–10 (2012)
11. A.B. Conn, N.I.M. Gould, P.L. Toint. *Trust Region Methods* (SIAM, Philadelphia, 2000)
12. J.E. Dennis, R.B. Schnabel, *Numerical Methods for Unconstrained Optimization and Nonlinear Equations*, 2nd edn. (SIAM, Philadelphia, 1996)
13. P. Dvurechensky, Yu. Nesterov, Global performance guarantees of second-order methods for unconstrained convex minimization, CORE Discussion Paper, 2018
14. A.V. Fiacco, G.P. McCormick, *Nonlinear Programming: Sequential Unconstrained Minimization Techniques* (Wiley, New York, 1968)
15. R. Fletcher, *Practical Methods of Optimization, Vol. 1, Unconstrained Minimization* (Wiley, New York, 1980)
16. M. Frank, P. Wolfe, An algorithm for quadratic programming. Nav. Res. Logist. Q. **3**, 149–154 (1956)
17. R.M. Freund, P. Grigas, New analysis and results for the Frank–Wolfe method. Math. Program. **155**, 199–230 (2014). https://doi.org/10.1007/s10107-014-0841-6
18. D. Garber, E. Hazan, A linearly convergent conditional gradient algorithm with application to online and stochastic optimization. arXiv: 1301.4666v5 (2013)

Y. Nesterov, *Lectures on Convex Optimization*, Springer Optimization and Its Applications 137, https://doi.org/10.1007/978-3-319-91578-4

19. D. Garber, E. Hazan, Faster rates for the Frank–Wolfe method over strongly convex sets. arXiv:1406.1305v2 (2015)
20. A. Gasnikov, Yu. Nesterov, Universal method for problems of stochastic composite minimization. Comput. Math. Math. Phys. **58**(1), 48–64 (2018)
21. S. Goldfeld, R. Quandt, H. Trotter, Maximization by quadratic hill climbing. Econometrica **34**, 541–551 (1966)
22. A. Griewank, The modification of Newton's method for unconstrained optimization by bounding cubic terms, Technical Report NA/12 (1981), Department of Applied Mathematics and Theoretical Physics, University of Cambridge, United Kingdom, 1981
23. Z. Harchaoui, A. Juditsky, A. Nemirovski, Conditional gradient algorithms for norm-regularized smooth convex optimization. Math. Program. **152**, 75–112 (2014). https://doi.org/10.1007/s10107-014-0778-9
24. J.-B. Hiriart-Urruty, C. Lemarechal, *Convex Analysis and Minimization Algorithms. Part 1*. A Series of Comprehensive Studies in Mathematics (Springer, Berlin, 1993)
25. M. Jaggi, Revisiting Frank–Wolfe: projection-free sparse convex optimization, in *Proceedings of the 30th International Conference on Machine Learning*, Atlanta, Georgia (2013)
26. L.V. Kantorovich, Functional analysis and applied mathematics. Uspehi Mat. Nauk **3**(1), 89–185 (1948) (in Russian). Translated as N.B.S. Report 1509, Washington D.C., 1952
27. L.G. Khachiyan, Rounding of polytopes in the real number model of computation. Math. Oper. Res. **21**(2), 307–320 (1996)
28. C. Lemarechal, A. Nemirovskii, Yu. Nesterov, New variants of bundle methods. Math. Program. **69**, 111–148 (1995)
29. K. Levenberg. A method for the solution of certain problems in least squares. Q. Appl. Math. **2**, 164–168 (1944)
30. D.G. Luenberger, *Linear and Nonlinear Programming*, 2nd edn. (Addison Wesley, Boston, 1984)
31. D. Marquardt, An algorithm for least-squares estimation of nonlinear parameters. SIAM J. Appl. Math. **11**, 431–441 (1963)
32. R. Monteiro, B. Svaiter, An accelerated hybrid proximal extragradient method for convex optimization and its implications to second-order methods. SIAM J. Optim. **23**(2), 1092–1125 (2013)
33. A. Nemirovski, Interior-point polynomial-time methods in convex programming (1996), https://www2.isye.gatech.edu/~nemirovs/LectIPM.pdf
34. A.S. Nemirovskij, D.B. Yudin, *Problem Complexity and Method Efficiency in Optimization*. Wiley-Interscience Series in Discrete Mathematics (A Wiley-Interscience Publication/Wiley, New York, 1983)
35. Yu. Nesterov, A method for unconstrained convex minimization problem with the rate of convergence $O(\frac{1}{k^2})$. Doklady AN SSSR **269**, 543–547 (1983) (In Russian; translated as Soviet Math. Docl.)
36. Yu. Nesterov, Minimization methods for nonsmooth convex and quasiconvex functions. *Ekonomika i Mat. Metody* **11**(3), 519–531 (1984) (In Russian; translated in *MatEcon*.)
37. Yu. Nesterov, *Efficient Methods in Nonlinear Programming* (Radio i Sviaz, Moscow, 1989) (In Russian.)
38. Yu. Nesterov, Semidefinite relaxation and nonconvex quadratic optimization. Optim. Methods Softw. **9**, 141–160 (1998)
39. Yu. Nesterov, *Introductory Lectures on Convex Optimization. A Basic Course* (Kluwer, Boston, 2004)
40. Yu. Nesterov, Smooth minimization of non-smooth functions. Math. Program. (A) **103**(1), 127–152 (2005)
41. Yu. Nesterov, Excessive gap technique in non-smooth convex minimizarion. SIAM J. Optim. **16** (1), 235–249 (2005)
42. Yu. Nesterov, Smoothing technique and its applications in semidefinite optimization. Math. Program. **110**(2), 245–259 (2007)

43. Yu. Nesterov, Modified Gauss–Newton scheme with worst-case guarantees for its global performance. Optim. Methods Softw. **22**(3), 469–483 (2007)
44. Yu. Nesterov, Rounding of convex sets and efficient gradient methods for linear programming problems. Optim. Methods Softw. **23**(1), 109–128 (2008)
45. Yu. Nesterov, Accelerating the cubic regularization of Newton's method on convex problems. Math. Program. **112**(1), 159–181 (2008)
46. Yu. Nesterov, Unconstrained convex minimization in relative scale. Math. Oper. Res. **34**(1), 180–193 (2009)
47. Yu. Nesterov, Barrier subgradient method. Math. Program. **127**(1), 31–56 (2011)
48. Yu. Nesterov, How to make the gradients small. Optima **88**, 10–11 (2012)
49. Yu. Nesterov, Towards non-symmetric conic optimization. Optim. Methods Softw. **27**(4–5), 893–918 (2012)
50. Yu. Nesterov, Complexity bounds for primal-dual methods minimizing the model of objective function. Math. Program. (2017). https://doi.org/10.1007/s10107-017-1188-6
51. Yu. Nesterov, A. Nemirovskii, *Interior-Point Polynomial Algorithms in Convex Programming* (SIAM, Philadelphia, 1994)
52. Yu. Nesterov, B. Polyak, Cubic regularization of Newton's method and its global performance. Math. Program. **108**(1), 177–205 (2006)
53. J. Nocedal, S.J. Wright, *Numerical Optimization* (Springer, New York, 1999)
54. J. Ortega, W. Rheinboldt, *Iterative Solution of Nonlinear Equations in Several Variables* (Academic Press, New York, 1970)
55. B.T. Polyak, *Introduction to Optimization* (Optimization Software, Publications Division, New York, 1987)
56. B.T. Polyak, History of mathematical programming in the USSR: analyzing the phenomenon. Math. Program. **91**(3), 401–416 (2002)
57. J. Renegar, *A Mathematical View of Interior-Point Methods in Convex Optimization*. MPS-SIAM Series on Optimization (SIAM, Philadelphia, 2001)
58. R.T. Rockafellar, *Convex Analysis* (Princeton University Press, Princeton, 1970)
59. C. Roos, T. Terlaky, J.-Ph. Vial, *Theory and Algorithms for Linear Optimization: An Interior Point Approach* (Wiley, Chichester, 1997)
60. M. Todd, *Minimum-Volume Ellipsoids: Theory and Algorithms*. MOS-SIAM Series on Optimization (SIAM, philadelphia, 2016)
61. M.J. Todd, E.A.Yildirim, On Khachiyan's algorithm for the computation of minimum volume enclosing ellipsoids, Technical Report, TR 1435, School of Operations Research and Industrial Engineering, Cornell University, 2005
62. S.J. Wright, *Primal-Dual Interior Point Methods* (SIAM, Philadelphia, 1996)
63. Y. Ye, *Interior Point Algorithms: Theory and Analysis* (Wiley, Hoboken, 1997)

Index

Y. Nesterov, *Lectures on Convex Optimization*, Springer Optimization
and Its Applications 137, https://doi.org/10.1007/978-3-319-91578-4