Duality in quasi–Newton methods and new variational characterizations of the DFP and BFGS updates

Osman Güler, Filiz Gürtuna and Olena Shevchenko*

September 2007

Abstract

It is known that quasi–Newton updates can be characterized by variational means, sometimes in more than one way. This paper has two main goals. We first formulate variational problems appearing in quasi-Newton methods within the vector space of symmetric matrices. This simplifies both their formulations and their subsequent solutions. We then construct, for the first time, duals of the variational problems for the DFP and BFGS updates and discover that the solution to a dual problem is either the same as the corresponding primal solution or the solutions are inverses of each other. Consequently, we obtain six new variational characterizations for the DFP and BFGS updates, three for each one.

Key words. Quasi–Newton, DFP, BFGS, variational problems, duality.
Abbreviated title: Duality and variational problems in quasi–Newton methods
AMS(MOS) subject classifications: primary: 90C53, 90C30, 90C46, 65K10, 49N15; secondary: 65K05, 49M37, 52A41.

*Department of Mathematics and Statistics, University of Maryland Baltimore County, Baltimore, Maryland 21250, USA. E-mail: {*guler,gurtuna1,olenshe1*}*@math.umbc.edu*. Research partially supported by the National Science Foundation under grant DMS–0411955

1 Introduction

The most successful and popular quasi–Newton method for unconstrained function minimization is the BFGS method of Broyden, Fletcher, Goldfarb, and Shanno [1, 7, 11, 15], followed by the DFP method of Davidon, Fletcher, and Powell [4, 10]. For easy reference, the update formulas for these two methods are presented below. In what follows, B–variables and H–variables represent Hessian and inverse Hessian approximations, respectively. The superscripts DFP and BFGS refer to the corresponding updates.

$$B_{k+1}^{\mathrm{DFP}} := (I - \gamma_k y_k s_k^T) B_k (I - \gamma_k s_k y_k^T) + \gamma_k y_k y_k^T, \tag{1.1}$$

$$B_{k+1}^{\mathrm{BFGS}} := B_k - \frac{B_k s_k s_k^T B_k}{s_k^T B_k s_k} + \gamma_k y_k y_k^T, \tag{1.2}$$

$$H_{k+1}^{\mathrm{DFP}} := H_k - \frac{H_k y_k y_k^T H_k}{y_k^T H_k y_k} + \gamma_k s_k s_k^T, \tag{1.3}$$

$$H_{k+1}^{\mathrm{BFGS}} := (I - \gamma_k s_k y_k^T) H_k (I - \gamma_k y_k s_k^T) + \gamma_k s_k s_k^T. \tag{1.4}$$

Here B_k and H_k are symmetric, positive definite $n \times n$ matrices and

$$y_k = \nabla f(x_{k+1}) - \nabla f(x_k), \quad s_k = x_{k+1} - x_k, \quad \text{and} \quad \gamma_k = \frac{1}{y_k^T s_k}, \tag{1.5}$$

where f is the function we would like to minimize, and x_k and x_{k+1} are approximations to a (local) minimizer of f.

The reader is referred to the survey papers of Dennis and Moré [5] and Nocedal [14] for more information on quasi–Newton methods. Although quasi–Newton methods were not originally discovered by variational means, it has been known since the early 1970s [12], [11] that the update formulas can be given a variational interpretation. For example, H_{k+1}^{BFGS} is the solution to the optimization problem

$$\begin{array}{ll} \min & \operatorname{tr}((H - H_k)W(H - H_k)W) \\ \text{s. t.} & H y_k = s_k, \\ & H^T = H, \end{array} \tag{1.6}$$

in which the decision variable H is an $n \times n$ matrix and where W is *any* symmetric positive definite matrix satisfying $W s_k = y_k$. It has been argued that it is desirable to impose the "secant condition" $H y_k = s_k$ and the symmetry condition $H^T = H$ on H_{k+1}, see [5]. Since there exist infinitely many symmetric matrices H satisfying the constraints, one should therefore choose the updated matrix H_{k+1} "close" to the current matrix H_k so as not to lose any information present in H_k. Using a measure of closeness such as the trace objective function above, we can obtain a unique approximation H_{k+1} to the Hessian of f at x_{k+1}. Different measures lead to different updates.

More variational characterizations of quasi–Newton updates appear later on. Byrd and Nocedal [2] use the function

$$\psi(X) = \operatorname{tr} X - \ln \det X$$

defined on symmetric, positive definite matrices, in their convergence analysis of the BFGS update. Subsequently, Fletcher [8] shows that the DFP and BFGS updates can be obtained from optimization problems involving the function $\psi(X)$. For example, H_{k+1}^{DFP} is the solution to the minimization problem

$$\min \left\{ \psi(B_k^{1/2} H B_k^{1/2}) : H y_k = s_k, H^T = H \right\}.$$

This paper has two goals. Our first goal is to establish a framework, within which we formulate variational problems appearing in quasi-Newton methods for unconstrained minimization. Our basic idea is simply to work directly in the vector space $\mathbb{SR}^{n \times n}$ of $n \times n$ symmetric matrices. We thereby handle the symmetry of our decision variables (the approximate Hessian matrices or their inverses) *implicitly*. Thus we take $\mathbb{SR}^{n \times n}$ as our universal vector space, venturing into the larger vector space $\mathbb{R}^{n \times n}$ only occasionally, in order to rewrite the secant equations in an appropriate form within $\mathbb{SR}^{n \times n}$. The desirable feature of this idea is that it *simplifies* the formulation of *most* of the variational problems we have encountered in quasi–Newton methods, and subsequently, their solutions. Our second goal is to investigate the *duals* of variational problems in quasi–Newton methods. We specifically formulate duals of several well known variational problems for the DFP and BFGS updates and discover that each primal–dual pair of problems has the remarkable feature that either the primal and dual solutions are the *same*, or they are *inverses* of each other. This is a situation that rarely happens in duality theory – usually primal and dual solutions are not related. Consequently, we obtain several *new* variational interpretations for the DFP and BFGS updates. Our two goals are related: we note that there is no unique or canonical formulation of dual problems. Using geometric language and working in the space of symmetric matrices $\mathbb{SR}^{n \times n}$ both contribute to the formulation of our dual problems that are simpler, more natural, and have the above mentioned properties. A straight forward formulation of the dual problem to (1.6) by writing down its Lagrangian function, for example, would lead to a dual problem which has two sets of decision variables (the multipliers corresponding to the constraints $H y_k = s_k$ and $H^T = H$), neither of which is a symmetric matrix.

The paper is organized as follows. In §2, we formulate the well known variational problems for the DFP and BFGS updates as least squares problems in $\mathbb{SR}^{n \times n}$. We then solve our least squares problems using a geometric language, avoiding Lagrange multipliers. Dennis and Schnabel [6] seem to be the first to give geometric solutions to the same problems. They first obtain a geometric solution to the least squares problem in $\mathbb{R}^{n \times n}$, relaxing the symmetry constraint on the decision variable, and then use the method of alternating projections to gain symmetry. The same paper also contains proofs of some of the results in this paper, such as Theorem 2.2, Corollary 2.3, Corollary 2.5, and Theorem 4.1, with different proofs than ours. Subsequently, Griewank [13] gives direct geometric proofs of the quasi–Newton updates similar to ours. In §3, we formulate and provide short solutions for the variational problems in Fletcher [8] involving the measure function $\psi(X)$ of Byrd and Nocedal [2]. In §4, we give short proofs for two of the variational results for sparse problems, one in Toint [16] and the other in Fletcher [9]. Our duality results are treated in the rest of the paper. In §5–§7, we provide a total of *six* new variational characterizations for the DFP and BFGS updates, three for each one. In §5, we formulate duals of the least squares problems from §2 and show that each primal–dual pair of problems have the *same* solution. This fact is traced in Theorem 5.1 to a remarkable property of geometric least squares problems that deserves to be better known. In §6, we give another dualization scheme for the least squares problems from §2, more in the spirit of approximation theory. They allow for a different interpretation of the DFP and BFGS updates. In §7, we formulate and solve the duals of the variational problems of §3. We discover that the solutions to each primal–dual pair of problems are *inverses* of each other. In some sense, the dual problems in this section have an advantage over the primal ones, since only one kind of update matrix (DFP or BFGS) appears (once as a variable and once as a solution) in each dual problem, in contrast to corresponding primal problem in which both DFP and BFGS updates appear. In the Appendix, we gather some results used in the main body of the paper.

Our notation is fairly standard. We use the inner product $\langle u, v\rangle = u^T v$ in $\mathbb{R}^n$ and the trace inner product

$$\langle X, Y\rangle = \operatorname{tr}(X^T Y) = \sum_{i,j=1}^{n} X_{ij} Y_{ij}$$

in the space $\mathbb{R}^{n\times n}$ of $n \times n$ matrices (hence in $\mathbb{SR}^{n\times n}$, the vector space of symmetric $n \times n$ matrices). If both inner products are used within the same formula, the meaning of each one should be clear from the context. We use several weighted trace inner products which are defined in the main body of the paper. The set of all symmetric $n \times n$ positive definite matrices will be denoted by $\mathbb{SR}^{n\times n}_{++}$. Let u be a vector and L be a linear subspace of a Euclidean vector space E. We denote by $\Pi_L u$ the orthogonal projection of u onto L, and $L^\perp$ denotes the orthogonal complement of L in E, that is, $L^\perp = \{v \in E : \langle u, v\rangle = 0,\ \forall u \in L\}$.

2 Least squares problems in quasi–Newton methods

In this section, we formulate some of the well known least squares problems appearing in quasi–Newton methods as variational problems in $\mathbb{SR}^{n\times n}$, and use a geometric approach to solve them.

We first need a preliminary result which is interesting in its own right.

Lemma 2.1. *Let s and y be vectors in $\mathbb{R}^n$, $s \neq 0$. The linear subspace corresponding to the affine subspace $\mathcal{A} := \{X \in \mathbb{SR}^{n\times n} : Xs = y\}$ is $\mathcal{L} = \{X \in \mathbb{SR}^{n\times n} : Xs = 0\}$. Let $\{u_i\}_1^n$ be a basis of $\mathbb{R}^n$, and define the matrices $S_i = su_i^T + u_i s^T$, $i = 1, \ldots, n$. The matrices $\{S_i\}_1^n$ are linearly independent and $\mathcal{L}$ is the intersection of n hyperplanes in $\mathbb{SR}^{n\times n}$,*

$$\mathcal{L} = \{X \in \mathbb{SR}^{n\times n} : \langle X, S_i\rangle = 0, \quad i = 1, \ldots, n\}.$$

Moreover,

$$\mathcal{L}^\perp = \operatorname{span}\{S_1, \ldots, S_n\} = \{s\lambda^T + \lambda s^T : \lambda \in \mathbb{R}^n\}. \tag{2.1}$$

Proof. The formula for $\mathcal{L}$ is obvious. Notice that the equation $Xs = 0$ in $\mathbb{R}^{n\times n}$ is equivalently given by its component equations $\langle X, su_i^T\rangle = 0$, $i = 1, \ldots, n$, since

$$0 = \langle u_i, Xs\rangle = \operatorname{tr}\left(u_i^T Xs\right) = \operatorname{tr}(Xsu_i^T) = \langle X, su_i^T\rangle, \quad i = 1, \ldots, n.$$

As X is symmetric, we also have

$$\langle X, u_i s^T\rangle = \operatorname{tr}(Xu_i s^T) = \operatorname{tr}(su_i^T X) = \operatorname{tr}(u_i^T Xs) = \operatorname{tr}(Xsu_i^T) = \langle X, su_i^T\rangle.$$

Thus, the equation $Xs = 0$ is equivalent to the equations $\langle X, u_i s^T\rangle = 0$, $i = 1, \ldots, n$. Consequently, $\mathcal{L} \subseteq \mathbb{SR}^{n\times n}$ can be written as an intersection of n hyperplanes $\langle X, S_i\rangle = 0$, $i = 1, \ldots, n$.

The formula $\mathcal{L}^\perp = \operatorname{span}\{S_1, \ldots, S_n\}$ follows immediately. Any linear combination $\sum_{i=1}^n \delta_i S_i \in \mathcal{L}^\perp$ can be written as

$$\sum_{i=1}^{n} \delta_i (su_i^T + u_i s^T) = s\lambda^T + \lambda s^T,$$

where $\lambda = \sum_{i=1}^n \delta_i u_i$.

The matrices $\{S_i\}_1^n$ are linearly independent: the equation $0 = \sum_1^n \delta_i S_i = \lambda s^T + s\lambda^T$ gives $0 = (s^T\lambda)\lambda + ||\lambda||^2 s$, and taking the inner product of both sides with s yields $(s^T\lambda)^2 + ||\lambda||^2 \cdot ||s||^2 = 0$. Thus, $||\lambda||^2 \cdot ||s||^2 = 0$, and since $s \neq 0$, we have $\lambda = 0$. Since $\lambda = \sum_{i=1}^n \delta_i u_i = 0$ and $\{u_i\}_1^n$ is a basis of $\mathbb{R}^n$, we have $\delta_i = 0$, $i = 1, \ldots, n$. □

We now consider a generic least squares problem, which is closely related to the variational problems having the DFP and BFGS updates as their solutions.

Theorem 2.2. (***Dennis and Schnabel [6]***) *The solution $\bar{X}$ to the problem in the vector space $\mathbb{SR}^{n\times n}$,*

$$\begin{aligned} \min \quad & \frac{1}{2}||X||^2 \\ \text{s.t.} \quad & Xs = y, \end{aligned} \tag{2.2}$$

is given by

$$\bar{X} = \frac{sy^T + ys^T}{\langle s, s\rangle} - \frac{\langle y, s\rangle}{\langle s, s\rangle^2} ss^T. \tag{2.3}$$

Proof. Define $f(X) = ||X||^2/2 = \langle X, X\rangle/2$. We have $\nabla f(X) = X$ and $\nabla^2 f(X) = I$, and the function f is convex. It follows from Lemma 8.1 in the Appendix that the solution $\bar{X}$ is characterized by the condition $\nabla f(\bar{X}) = \bar{X} \in \mathcal{L}^\perp$. Consequently, Lemma 2.1 implies that $\bar{X}$ is characterized by the equation

$$\bar{X} = \lambda s^T + s\lambda^T$$

for some $\lambda \in \mathbb{R}^n$. We have

$$\langle y, s\rangle = \langle \bar{X}s, s\rangle = \langle [\lambda s^T + s\lambda^T]s, s\rangle = 2\langle \lambda, s\rangle ||s||^2,$$

and $\langle \lambda, s\rangle = \langle y, s\rangle/(2||s||^2)$. Substituting this in the equation $y = \bar{X}s = \langle \lambda, s\rangle s + ||s||^2\lambda$ gives

$$\lambda = \frac{1}{||s||^2} y - \frac{\langle y, s\rangle}{2||s||^4} s.$$

Finally, substituting this in $\bar{X} = \lambda s^T + s\lambda^T$ gives (2.3). □

Corollary 2.3. (***Dennis and Schnabel [6]***) *Let $X_0 \in \mathbb{SR}^{n\times n}$ and $W \in \mathbb{SR}^{n\times n}_{++}$ be a weighting matrix. The solution $\bar{X}$ to the problem in the vector space $\mathbb{SR}^{n\times n}$,*

$$\begin{aligned} \min \quad & \frac{1}{2}\|W^{1/2}(X - X_0)W^{1/2}\|^2 \\ \text{s.t.} \quad & Xs = y, \end{aligned} \tag{2.4}$$

is given by

$$\begin{aligned} \bar{X} = & X_0 + \frac{W^{-1}s(y - X_0 s)^T + (y - X_0 s)s^T W^{-1}}{\langle s, W^{-1}s\rangle} \\ & - \frac{\langle y - X_0 s, s\rangle}{\langle s, W^{-1}s\rangle^2} W^{-1}ss^T W^{-1}. \end{aligned}$$

Proof. With the following change of variables

$$\tilde{X} := W^{1/2}(X - X_0)W^{1/2}, \quad \tilde{y} := W^{1/2}(y - X_0 s), \quad \tilde{s} := W^{-1/2}s,$$

the problem (2.4) reduces to problem (2.2). After substituting the expressions for $\tilde{X}$, $\tilde{y}$, and $\tilde{s}$ into (2.3), we multiply the resulting equality by $W^{-1/2}$ from both sides to get the desired expression for $\bar{X}$. □

Now, the DFP and BFGS updates follow from Corollary 2.3.

Corollary 2.4. *The update matrix* B_{k+1}^{DFP} *in* (1.1) *is the solution to the problem*

$$\begin{aligned} \min \quad & \frac{1}{2}\|W^{1/2}(B - B_k)W^{1/2}\|^2 \\ \text{s.t.} \quad & Bs_k = y_k, \end{aligned} \tag{2.5}$$

where $B_k \in \mathbb{SR}^{n\times n}_{++}$ *and* $W \in \mathbb{SR}^{n\times n}_{++}$ *is any matrix satisfying* $Wy_k = s_k$.

Proof. Using Corollary 2.3, we have

$$\begin{aligned} B_{k+1} =& B_k + \frac{W^{-1}s_k(y_k - B_k s_k)^T + (y_k - B_k s_k)s_k^T W^{-1}}{\langle s_k, W^{-1}s_k\rangle} \\ & - \frac{\langle y_k - B_k s_k, s_k\rangle}{\langle s_k, W^{-1}s_k\rangle^2} W^{-1}s_k s_k^T W^{-1}. \end{aligned}$$

The requirement $Wy_k = s_k$ (or $y_k = W^{-1}s_k$) *simplifies* the above expression and makes B_{k+1} independent of W. It is a routine to verify that the resulting formula for B_{k+1} is the same as the one obtained by expanding (1.1). □

It is evident from (1.1) that $B_{k+1} = G^T G + F$ where $F = \gamma_k y_k y_k^T$ and $G = B_k^{1/2}(I - \gamma_k s_k y_k^T)$. Since both $G^T G$ and F are positive semidefinite matrices, so is B_{k+1}. Moreover, $Gd = 0$ for $d \neq 0$ if and only if d is a multiple of s_k, but then $\langle Fd, d\rangle > 0$. Thus, we see that if B_k is positive definite and $\langle s_k, y_k\rangle > 0$, then the matrix B_{k+1} is also positive definite.

Corollary 2.5. *The update matrix* H_{k+1}^{BFGS} *in* (1.4) *is the solution to the problem*

$$\begin{aligned} \min \quad & \frac{1}{2}\|W^{1/2}(H - H_k)W^{1/2}\|^2 \\ \text{s.t.} \quad & Hy_k = s_k, \end{aligned} \tag{2.6}$$

where $H_k \in \mathbb{SR}^{n\times n}_{++}$ *and* $W \in \mathbb{SR}^{n\times n}_{++}$ *is any matrix satisfying* $Ws_k = y_k$.

Proof. The proof is similar to the proof of Corollary 2.4. □

As in the DFP case above, we conclude that if H_k is positive definite and $\langle y_k, s_k\rangle > 0$, then H_{k+1} is positive definite.

3 Trace–determinant function minimization problems in quasi–Newton methods

In this section, we present short and more geometric proofs of the main results in Fletcher [8].

Theorem 3.1. *If the affine set $\{X \in \mathbb{SR}^{n\times n} : Xs = y\}$ contains a positive definite matrix, then the solution $\bar{X}$ to the problem*

$$\begin{aligned} \min \quad & \psi(X) = \langle I, X\rangle - \ln\det X \\ \text{s.t.} \quad & Xs = y \end{aligned} \tag{3.1}$$

in the vector space $\mathbb{SR}^{n\times n}$ satisfies

$$\bar{X}^{-1} = I + \frac{ss^T}{\langle y, s\rangle} - \frac{sy^T + ys^T}{\langle y, s\rangle} + \frac{\langle y, y\rangle}{\langle y, s\rangle^2} ss^T. \tag{3.2}$$

Proof. The gradient and the Hessian of ψ are given by

$$\nabla\psi(x) = I - X^{-1}, \quad \nabla^2\psi(x) = X^{-1} \otimes X^{-1}.$$

See equation (8.2) in the Appendix. Thus, ψ is strictly convex on the cone of positive definite matrices in $\mathbb{SR}^{n\times n}$. It is also coercive on the same cone. Lemma 8.1 and Lemma 2.1 imply that the solution $\bar{X}$ satisfies the condition

$$I - \bar{X}^{-1} = s\lambda^T + \lambda s^T \tag{3.3}$$

for some $\lambda \in \mathbb{R}^n$. The secant equation $Xs = y$ gives $\bar{X}^{-1}y = s$, and we have

$$\langle y - s, y\rangle = \langle (I - \bar{X}^{-1})y, y\rangle = \langle [s\lambda^T + \lambda s^T]y, y\rangle = 2\langle \lambda, y\rangle \, \langle y, s\rangle,$$

which yields $\langle \lambda, y\rangle = \langle y - s, y\rangle / (2\langle y, s\rangle)$. Then substituting this in the equation

$$y - s = (I - \bar{X}^{-1})y = \langle \lambda, y\rangle s + \langle y, s\rangle \lambda$$

gives

$$\lambda = \frac{y - s}{\langle s, y\rangle} + \frac{\langle s - y, y\rangle}{2\langle s, y\rangle^2} s.$$

Then substituting this in (3.3) and simplifying the result yields (3.2). □

Corollary 3.2. *Let $H_k \in \mathbb{SR}^{n\times n}$ be a positive definite matrix and assume that $\langle s_k, y_k\rangle > 0$. The update matrix H_{k+1}^{BFGS} in (1.4) satisfies $H_{k+1} = \bar{B}^{-1}$, where $\bar{B}$ is the solution to the problem*

$$\begin{aligned} \min \quad & \psi(H_k^{1/2} B H_k^{1/2}) = \langle H_k, B\rangle - \ln\det B + const \\ \text{s.t.} \quad & Bs_k = y_k. \end{aligned} \tag{3.4}$$

Proof. The change of variables $X = H_k^{1/2} B H_k^{1/2}$, $y = H_k^{1/2} y_k$, and $s = H_k^{-1/2} s_k$ reduces the problem to Problem (3.1) in Theorem 3.1. Substituting the values of X, y, s above in equation (3.2) and simplifying, we obtain

$$\bar{B}^{-1} = H_k - \frac{s_k y_k^T H_k + H_k y_k s_k^T}{\langle s_k, y_k\rangle} + \frac{s_k s_k^T}{\langle s_k, y_k\rangle} + \frac{\langle H_k y_k, y_k\rangle}{\langle s_k, y_k\rangle^2} s_k s_k^T.$$

The right-hand side of this formula is identical to (1.4). Consequently, $\bar{B} = H_{k+1}^{-1}$. □

Corollary 3.3. *Let $B_k \in \mathbb{SR}^{n\times n}$ be a positive definite matrix and assume that $\langle s_k, y_k\rangle > 0$. The update matrix B_{k+1}^{DFP} in* (1.1) *satisfies $B_{k+1} = \bar{H}^{-1}$, where $\bar{H}$ is the solution to the problem*

$$\begin{array}{ll} \min & \psi(B_k^{1/2} H B_k^{1/2}) = \langle B_k, H\rangle - \ln\det H + const \\ \text{s.t.} & H y_k = s_k. \end{array}$$

Proof. This is similar to the proof of Corollary 3.2, using the change of variables $X = B_k^{1/2} H B_k^{1/2}$, $y = B_k^{1/2} s_k$, and $s = B_k^{-1/2} y_k$. □

4 Variational problems arising in sparse quasi–Newton methods

In this section, we give short solutions to two variational problems, one in Toint [16], and the other in Fletcher [9].

Theorem 4.1. **(*Toint [16]*)** *Let $S \subset \{(i,j) : 1 \le i,j \le n\}$. Consider the minimization problem in the vector space $\mathbb{SR}^{n\times n}$,*

$$\begin{array}{ll} \min & \frac{1}{2}||X||^2 \\ \text{s.t.} & Xs = y, \\ & X_{ij} = 0, \quad (i,j) \in S. \end{array} \tag{4.1}$$

Define $\mathcal{L} := \{X : X_{ij} = 0, (i,j) \in S\}$. The solution to (4.1) *is given by*

$$\bar{X} = \Pi_{\mathcal{L}}(\lambda s^T + s\lambda^T), \tag{4.2}$$

where λ is the solution of the linear equations $Q\lambda = y$ in $\mathbb{R}^n$, and where $Q^T = [s_1, s_2, \ldots, s_n]$, $s_i = S_i s$, and $S_i = \Pi_{\mathcal{L}}(s e_i^T + e_i s^T)$, $i = 1, \ldots, n$.

Proof. Define $\mathcal{M} := \{X : Xs = 0\}$. Lemma 8.1 implies that $\bar{X} \in (\mathcal{M} \cap \mathcal{L})^\perp = \mathcal{M}^\perp + \mathcal{L}^\perp$. Write

$$\bar{X} = (\lambda s^T + s\lambda^T) + \Lambda$$

with $\lambda s^T + s\lambda^T \in \mathcal{M}^\perp$ (see Lemma 2.1), and $\Lambda \in \mathcal{L}^\perp$. Since $\bar{X} \in \mathcal{L}$, we see that

$$\bar{X} = \Pi_{\mathcal{L}}(\lambda s^T + s\lambda^T).$$

We have $y = \bar{X}s = \Pi_{\mathcal{L}}(\lambda s^T + s\lambda^T)s$, and

$$\begin{aligned} y_i &= \left\langle \left[\Pi_{\mathcal{L}}(\lambda s^T + s\lambda^T)\right] s, e_i \right\rangle = \left\langle \Pi_{\mathcal{L}}(\lambda s^T + s\lambda^T), \frac{e_i s^T + s e_i^T}{2} \right\rangle \\ &= \left\langle \frac{\lambda s^T + s\lambda^T}{2}, \Pi_{\mathcal{L}}(e_i s^T + s e_i^T) \right\rangle = \left\langle \frac{\lambda s^T + s\lambda^T}{2}, S_i \right\rangle \\ &= \langle S_i s, \lambda\rangle = \langle s_i, \lambda\rangle. \end{aligned}$$

□

The projection operator $\Pi_{\mathcal{L}}$ is the so–called "gangster operator" defined by

$$\mathcal{G}(H)_{ij} = \begin{cases} 0 & (i,j) \in S, \\ H_{ij} & (i,j) \in S^{\perp} \end{cases}$$

with $S^{\perp}$ denoting the complement of S, because it shoots "holes" at the entries $(i,j) \in S$ of matrix H.

We remark that the matrix Q is symmetric, since

$$Q_{ij} = (s_i)_j = \langle s_i, e_j \rangle = \langle \Pi_{\mathcal{L}}(se_i^T + e_i s^T)s, e_j \rangle = \left\langle \Pi_{\mathcal{L}}(se_i^T + e_i s^T), \frac{se_j^T + e_j s^T}{2} \right\rangle$$
$$= \left\langle \frac{se_i^T + e_i s^T}{2}, \Pi_{\mathcal{L}}(se_j^T + e_j s^T) \right\rangle = Q_{ji},$$

and has the same sparsity pattern S: we have

$$Q_{ij} = \left\langle \Pi_{\mathcal{L}}(se_i^T + e_i s^T), \frac{se_j^T + e_j s^T}{2} \right\rangle = \langle \Pi_{\mathcal{L}}(se_i^T + e_i s^T)e_j, s \rangle,$$

and it is easy to show that $\Pi_{\mathcal{L}}(se_i^T + e_i s^T)e_j = 0$ if $(i,j) \in S$.

Theorem 4.2. **(*Fletcher [9]*)** *Let $B_k \in \mathbb{SR}^{n\times n}$ be positive definite, and $H_k = B_k^{-1}$. The solution $\bar{B}$ to the minimization problem in the vector space $\mathbb{SR}^{n\times n}$,*

$$\begin{array}{ll} \min & \psi_{H_k}(B) := \langle H_k, B \rangle - \ln\det(B) \\ \text{s. t.} & Bs_k = y_k, \\ & B_{ij} = 0, \ (i,j) \in S, \end{array} \tag{4.3}$$

is characterized by the existence of λ such that

$$\mathcal{G}(\bar{H}) = \mathcal{G}(H_k + \lambda s^T + s\lambda^T),$$

where $\bar{H} = \bar{B}^{-1}$.

Proof. As in the proof of Theorem 4.1, define $\mathcal{L} := \{B : B_{ij} = 0, (i,j) \in S\}$ and $\mathcal{M} := \{B : Bs_k = 0\}$. The solution $\bar{B}$ to (4.3) satisfies

$$\nabla_B \psi_{H_k}(\bar{B}) = H_k - \bar{B}^{-1} \in \mathcal{M}^{\perp} + \mathcal{L}^{\perp},$$

that is,

$$\bar{B}^{-1} = H_k + \lambda s_k^T + s_k \lambda^T + \Lambda,$$

where $\Lambda \in \mathcal{L}^{\perp}$. The theorem is proved since $\mathcal{G}(\Lambda) = 0$. □

5 Dual least squares problems

Although the minimization problems (2.5) and (2.6) have been well known in the literature since the early 1970s, it seems that the associated dual problems have not been studied so far. In this section, we give dual problems for the least squares minimization problem.

We first consider a primal–dual pair of geometric least squares problems, see Courant–Hilbert [3], pp. 252–257.

Theorem 5.1. *Let x_0 and y_0 be points in a Euclidean space E, and L be a linear subspace of E. The least squares problems*

$$(P) \quad \min \quad \frac{1}{2}||x - x_0||^2 \qquad\qquad (D) \quad \min \quad \frac{1}{2}||y - y_0||^2$$
$$x \in y_0 + L, \qquad\qquad\qquad\qquad y \in x_0 + L^{\perp},$$

are duals of each other. Furthermore, they have the same *solution.*

Proof. Note that problem (P) can be written as the minimax problem

$$\min_{x \in E} \max_{\lambda \in L^{\perp}} L(x, \lambda) := \frac{1}{2}||x - x_0||^2 + \langle y_0 - x, \lambda \rangle,$$

since $\max_{\lambda \in L^{\perp}} \langle x - y_0, \lambda \rangle = 0$ if $x \in y_0 + L$, and $+\infty$ otherwise. The dual problem with respect to the Lagrangian function $L(x, \lambda)$ is the maximin problem

$$\max_{\lambda \in L^{\perp}} \min_{x \in E} L(x, \lambda) := \frac{1}{2}||x - x_0||^2 + \langle y_0 - x, \lambda \rangle.$$

The inner minimum is achieved at the point $x^* = x_0 + \lambda$. Substituting this in L and rearranging its terms, we obtain $L(x^*, \lambda) = -||\lambda + x_0 - y_0||^2/2 + ||x_0 - y_0||^2/2$. Thus the dual problem becomes, up to an additive constant $||x_0 - y_0||^2/2$,

$$\max_{\lambda \in L^{\perp}} -\frac{1}{2}||\lambda + x_0 - y_0||^2 = -\min_{\lambda \in L^{\perp}} \frac{1}{2}||\lambda + x_0 - y_0||^2.$$

With the change of variables $y = \lambda + x_0$, the right-hand side problem above is equivalent to (D).

Now, let x^* and y^* be the solutions to (P) and (D), respectively. We have

$$x^* - x_0 \in L^{\perp}, \quad x^* - y_0 \in L, \qquad y^* - y_0 \in L, \quad y^* - x_0 \in L^{\perp},$$

where the first and third inclusions follow from Lemma 8.1. These imply $x^* - y^* = (x^* - x_0) - (y^* - x_0) \in L^{\perp}$ and $x^* - y^* = (x^* - y_0) - (y^* - y_0) \in L$. Thus, $x^* - y^* \in L \cap L^{\perp} = \{0\}$, that is, $x^* = y^*$. □

Remark 5.2. We emphasize that the above pair of least squares problems (P) and (D) have the same solution. This is illustrated in Figure 1. Note that the primal problem (P) is the (orthogonal) projection of the point x_0 on the lower affine subspace $x_0 + L$ onto the upper affine subspace $y_0 + L$, whereas the dual problem (D) is the projection of the point y_0 on the upper affine subspace $x_0 + L$ onto the complementary affine subspace $x_0 + L^{\perp}$. As the proof above shows, the equality $x^* = y^*$ of the solutions, together with the Strong Duality Theorem (which holds true in this case since no constraint qualification is needed for affine constraints), implies the equality

$$\frac{1}{2}||x^* - x_0||^2 = -\frac{1}{2}||x^* - y_0||^2 + \frac{1}{2}||x_0 - y_0||^2,$$

where the left-hand side is the value of the minimax problem and the right-hand side is the value of the maximin problem. This amounts to the equation

$$||x_0 - y_0||^2 = ||x_0 - x^*||^2 + ||x^* - y_0||^2, \quad x_0 - x^* \in L^{\perp}, x^* - y_0 \in L,$$

which is precisely the Pythagorean theorem applied to the triangle with vertices $\{x_0, y_0, x^*\}$.

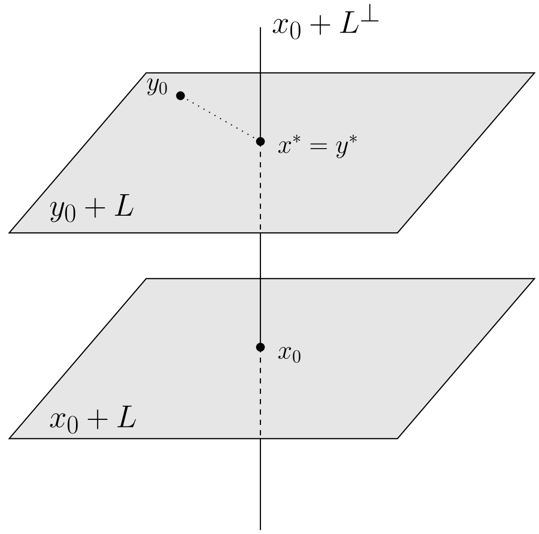

Figure 1: Illustration of Theorem 5.1

We now use Theorem 5.1 to obtain the duals of the least squares problems (2.5) and (2.6). By Theorem 5.1, these dual problems give *new* variational characterizations of the DFP and BFGS updates.

Note that in problem (2.5), if we make the change of variables

$$\tilde{B} := W^{1/2}BW^{1/2}, \quad \tilde{B}_k := W^{1/2}B_kW^{1/2}, \quad \tilde{y}_k := W^{1/2}y_k, \quad \tilde{s}_k := W^{-1/2}s_k,$$

we arrive at the least squares problem

$$\begin{array}{ll} \min & \frac{1}{2}||\tilde{B} - \tilde{B}_k||^2 \\ \text{s.t.} & \tilde{B}\,\tilde{s}_k = \tilde{y}_k. \end{array}$$

Theorem 5.1 and Lemma 2.1 can then be used to obtain a dual least squares problem, which in turn can be transformed into another least squares problem in terms of the original variables. Equivalently, we can obtain this dual problem in a different fashion, by changing the inner product instead of the variables: let $W \in \mathbb{SR}^{n\times n}$ be a positive definite matrix. Consider W–norm on $\mathbb{SR}^{n\times n}$ given by

$$||X||_W^2 := \operatorname{tr}(W^{1/2}XW^{1/2})^2 = \operatorname{tr}(WXWX),$$

and the corresponding inner–product

$$\langle X, Y\rangle_W := \operatorname{tr}(WXWY) = \operatorname{tr}((W \otimes W)XY) = \langle (W \otimes W)X, Y\rangle,$$

where

$$(W \otimes W)X := WXW.$$

In the Euclidean space $(\mathbb{SR}^{n\times n}, ||\cdot||_W)$, the problem (2.5) becomes

$$\begin{array}{ll} \min & \frac{1}{2}||B - B_k||_W^2 \\ \text{s.t.} & Bs_k = y_k, \end{array} \tag{5.1}$$

to which Theorem 5.1 applies. Let $\overline{B}$ be any matrix in the affine constraint set $\mathcal{A} := \{B \in \mathbb{SR}^{n\times n} : Bs_k = y_k\}$. Then $\mathcal{A} = \overline{B} + \mathcal{L}$ where $\mathcal{L} = \{B : Bs_k = 0\}$. In order to determine the dual problem in this setting, we need to compute the orthogonal complement of $\mathcal{L}$. This is done in the lemma below, which is an analogue of Lemma 2.1.

Lemma 5.3. *Let $W \in \mathbb{SR}^{n\times n}$ be a positive definite matrix and $s \in \mathbb{R}^n$ be a nonzero vector. The orthogonal complement of the linear subspace $\mathcal{L} = \{B : Bs = 0\}$ in the Euclidean space $(\mathbb{SR}^{n\times n}, ||\cdot||_W)$ is*

$$\mathcal{L}^\perp = \{\lambda(W^{-1}s)^T + (W^{-1}s)\lambda^T : \lambda \in \mathbb{R}^n\}.$$

Proof. Let $\{u_i\}_1^n$ be a basis of $\mathbb{R}^n$. $\mathcal{L}$ is characterized by the component equations $u_i^T Bs = s^T Bu_i = 0$, $i = 1,\dots,n$, or equivalently, by the equations

$$\begin{aligned} 0 &= \langle B, u_i s^T + s u_i^T\rangle = \langle B, W^{-1}(u_i s^T + s u_i^T)W^{-1}\rangle_W, \\ &= \langle B, (W^{-1}u_i)(W^{-1}s)^T + (W^{-1}s)(W^{-1}u_i)^T\rangle_W, \quad i = 1,\dots,n. \end{aligned}$$

It follows that

$$\begin{aligned} \mathcal{L}^\perp &= \operatorname{span}\{(W^{-1}u_i)(W^{-1}s)^T + (W^{-1}s)(W^{-1}u_i)^T : i = 1,\dots,n\}, \\ &= \{\lambda(W^{-1}s)^T + (W^{-1}s)\lambda^T : \lambda \in \mathbb{R}^n\}. \end{aligned}$$

□

Corollary 5.4. *The update matrix B_{k+1}^{DFP} in* (1.1) *is the solution to the least squares problem*

$$\min_{\lambda\in\mathbb{R}^n} \frac{1}{2}||B_k - \hat{B} + \lambda y_k^T + y_k\lambda^T||_W^2, \tag{5.2}$$

where y_k, s_k are defined in (1.5)*, $B_k \in \mathbb{SR}^{n\times n}_{++}$, and W satisfies the conditions in Corollary 2.4, and $\hat{B}$ is* any *matrix in $\mathbb{SR}^{n\times n}$ satisfying the secant equation $\hat{B}s_k = y_k$. In particular, we may choose*

$$\hat{B} = \frac{y_k y_k^T}{\langle s_k, y_k\rangle}.$$

Proof. The affine constraint set in (5.1) is $\hat{B} + \mathcal{L}$, where $\mathcal{L} = \{B : Bs_k = 0\}$, and we have $W^{-1}s_k = y_k$. The proof follows immediately from Theorem 5.1 and Lemma 5.3. □

Similarly, we have

Corollary 5.5. *The update matrix H_{k+1}^{BFGS} in* (1.4) *is the solution to the least squares problem*

$$\min_{\lambda\in\mathbb{R}^n} \frac{1}{2}||H_k - \hat{H} + \lambda s_k^T + s_k\lambda^T||_W^2, \tag{5.3}$$

where y_k, s_k are defined in (1.5)*, $B_k \in \mathbb{SR}^{n\times n}_{++}$, and W satisfies the conditions in Corollary 2.5, and $\hat{H}$ is any matrix in $\mathbb{SR}^{n\times n}$ satisfying the secant equation $\hat{H}y_k = s_k$. In particular, we may choose*

$$\hat{H} = \frac{s_k s_k^T}{\langle s_k, y_k\rangle}.$$

6 Another dualization of the least squares problems

We now give a different pair of dual problems for the minimization problems (2.5) and (2.6). They provide another interpretation of the DFP and BFGS updates.

Theorem 6.1. *Let W satisfy the conditions in Corollary 2.4, $\hat{B} \in \mathbb{SR}^{n\times n}$ be any matrix satisfying $\hat{B}s_k = y_k$ (a convenient choice is $\hat{B} = (y_k y_k^T)/\langle s_k, y_k\rangle$), and $B_k \in \mathbb{SR}^{n\times n}_{++}$. The problem*

$$\begin{array}{ll} \min & \langle B_k - \hat{B}, Y\rangle_W \\ \text{s. t.} & \|Y\|_W \le 1, \\ & Y = \lambda y_k^T + y_k \lambda^T, \lambda \in \mathbb{R}^n, \end{array} \tag{6.1}$$

is dual to problem (2.5) *and the DFP update matrix is given by*

$$B_{k+1}^{DFP} = B_k + \alpha \bar{Y},$$

where $\bar{Y}$ is the solution to (6.1) *and α is chosen so that the secant equation $B_{k+1}s_k = y_k$ is satisfied.*

Proof. We write problem (2.5) in the form $\min\{\|B - B_k\|_W : B \in \hat{B} + \mathcal{L}\}$, where $\mathcal{L} = \{B : Bs_k = 0\}$. We have

$$\begin{aligned} \min_{B\in\hat{B}+\mathcal{L}} \|B - B_k\|_W &= \min_{B\in\hat{B}+\mathcal{L}} \quad \max_{\|Y\|_W\le 1} \langle B - B_k, Y\rangle_W \\ &= \max_{\|Y\|_W\le 1} \quad \min_{B\in\hat{B}+\mathcal{L}} \langle B - B_k, Y\rangle_W \\ &= \max_{\|Y\|_W\le 1} \left\{ \langle \hat{B} - B_k, Y\rangle_W + \min_{X\in\mathcal{L}} \langle X, Y\rangle_W \right\} \\ &= \max_{\|Y\|_W\le 1, Y\in\mathcal{L}^\perp} \langle \hat{B} - B_k, Y\rangle_W, \end{aligned}$$

where the second equality follows from the minimax theorem, and the last equality follows from the fact that $\min\{\langle X, Y\rangle_W : X \in \mathcal{L}\}$ equals zero if $Y \in \mathcal{L}^\perp$, and $-\infty$ otherwise. Using the Cauchy–Schwarz inequality, the first equality above holds only if $B_{k+1} - B_k = \alpha\bar{Y}$ for some $\alpha \in \mathbb{R}$. This completes the proof, since $\mathcal{L}^\perp$ is given in Lemma 5.3 and $W^{-1}s_k = y_k$. □

We note that the dual solution $\bar{Y}$ is the point in $\mathcal{L}^\perp$ making the *smallest angle* (in the W–inner product) with the point $\hat{B} - B_k$.

Similarly, we have

Theorem 6.2. *Let W satisfy the conditions in Corollary 2.5, $\hat{H} \in \mathbb{SR}^{n\times n}$ be any matrix satisfying $\hat{H}y_k = s_k$ (a convenient choice is $\hat{H} = (s_k s_k^T)/\langle s_k, y_k\rangle$), and $H_k \in \mathbb{SR}^{n\times n}_{++}$. The problem*

$$\begin{array}{ll} \min & \langle H_k - \hat{H}, Z\rangle_W \\ \text{s. t.} & \|Z\|_W \le 1, \\ & Z = \lambda s_k^T + s_k \lambda^T, \lambda \in \mathbb{R}^n, \end{array} \tag{6.2}$$

is dual to problem (2.6) *and the BFGS update matrix is given by*

$$H_{k+1}^{BFGS} = H_k + \alpha \bar{Z},$$

where $\bar{Z}$ is the solution to (6.2) *and α is chosen so that the secant equation $H_{k+1}y_k = s_k$ is satisfied.*

7 Dual of the trace–determinant function minimization problem

In this section, we investigate the dual problem to the minimization problem (3.4) in §3, which does not seem to be studied in the literature. The primal–dual pair of problems have similar objective functions and related solutions. Consequently, we obtain additional variational characterizations for the DFP and BFGS updates.

We first consider a generic primal–dual pair of problems from which the DFP and BFGS updates follow as easy corollaries.

Theorem 7.1. *Let X_0 and Y_0 be matrices in $\mathbb{SR}^{n\times n}$. The following minimization problems are duals of each other,*

$$\begin{array}{llll} (P) & \min & \langle X_0, X\rangle - \ln\det X \\ & & X \in Y_0 + \mathcal{L}, \end{array} \qquad \begin{array}{llll} (D) & \min & \langle Y_0, Y\rangle - \ln\det Y \\ & & Y \in X_0 + \mathcal{L}^\perp. \end{array}$$

If both (P) and (D) have positive definite feasible solutions, then they both have (optimal) solutions and the Strong Duality Theorem holds. Furthermore, the solutions of (P) and (D) are inverses *of each other, that is, $\bar{Y} = (\bar{X})^{-1}$ where $\bar{X}$ and $\bar{Y}$ are the solutions of (P) and (D), respectively.*

Proof. Since the objective functions in (P) and (D) are coercive, both problems have solutions and the Strong Duality Theorem holds true. The primal problem (P) can be written as the minimax problem

$$\min_X \max_{Z\in\mathcal{L}^\perp} L(X,Z) := \langle X_0, X\rangle - \ln\det X + \langle X - Y_0, Z\rangle,$$

since $\max_{Z\in\mathcal{L}^\perp} \langle X - Y_0, Z\rangle = 0$ if $X - Y_0 \in \mathcal{L}$, and $+\infty$ otherwise. The dual problem with respect to the Lagrangian function $L(X,Z)$ is

$$\max_{Z\in\mathcal{L}^\perp} \min_X \left\{\langle X_0, X\rangle - \ln\det X + \langle X - Y_0, Z\rangle\right\}.$$

The inner minimum is achieved at the point $\tilde{X}$ satisfying

$$(\tilde{X})^{-1} = X_0 + Z. \tag{7.1}$$

Substituting this in $L(X,Z)$ and simplifying, we arrive at the equation $L(\tilde{X},Z) = n - \langle Y_0, Z\rangle + \ln\det(X_0+Z)$. Thus, the dual problem is equivalent to

$$\min_{Z\in\mathcal{L}^\perp} \left\{\langle Y_0, Z\rangle - \ln\det(X_0+Z)\right\}.$$

With the change of variables $Y = X_0 + Z$, and using the description of $\mathcal{L}^\perp$ in (2.1), we see that the above problem is equivalent to (D). It follows from (7.1) that $\bar{Y} = (\bar{X})^{-1}$. □

The new characterizations of the DFP and BFGS update formulas are immediate consequences of this theorem.

Corollary 7.2. *Let H_k be a symmetric positive definite matrix. The update matrix H_{k+1}^{BFGS} in* (1.4) *is the solution to the problem*

$$\begin{array}{ll} \min & \langle \hat{Y}, Y\rangle - \ln\det Y \\ & Y = H_k + \lambda s_k^T + s_k\lambda^T, \quad \lambda \in \mathbb{R}^n, \end{array} \tag{7.2}$$

where $\hat{Y}$ is any *matrix in $\mathbb{SR}^{n\times n}$ satisfying $\hat{Y}s_k = y_k$ (a convenient choice is $\hat{Y} = y_k y_k^T/\langle s_k, y_k\rangle$).*

Proof. Define $X_0 = H_k$ and $\mathcal{L} = \{Y \in \mathbb{SR}^{n\times n} : Ys_k = 0\}$. The proof follows from Theorem 7.1 and Corollary 3.2. □

Thus, we obtain here a new result that the BFGS update matrices B_{k+1} and $H_{k+1} := B_{k+1}^{-1}$ come from the primal-dual problems (3.4) and (7.2), respectively.

Similarly, we have

Corollary 7.3. *Let B_k be a symmetric positive definite matrix. The update matrix B_{k+1}^{DFP} in* (1.1) *is the solution to the problem*

$$\begin{array}{ll} \min & \langle \hat{Y}, Y\rangle - \ln\det Y \\ & Y = B_k + \lambda s_k^T + s_k\lambda^T, \quad \lambda \in \mathbb{R}^n, \end{array} \tag{7.3}$$

where $\hat{Y}$ is any *matrix in $\mathbb{SR}^{n\times n}$ satisfying $\hat{Y}y_k = s_k$ (such as $\hat{Y} = s_k s_k^T/\langle s_k, y_k\rangle$).*

8 Appendix

In this appendix, we collect for completeness some results used in the main body of the paper.

Lemma 8.1. *Let $A = a + L \subseteq E$ be an affine set in a Euclidean space E where L is a linear subspace of E. Let $f : A \to \mathbb{R}$ be a differentiable function, and consider the problem* $\min\{f(x) : x \in A\}$. *If $\bar{x} \in A$ is a local minimizer of f, then*

$$\nabla f(\bar{x}) \in L^\perp. \tag{8.1}$$

If f is convex, then (8.1) *is a sufficient condition for $\bar{x}$ to be a global minimizer of f on A.*

Proof. Let x be an arbitrary point of A. For $|t|$ small enough, we have

$$f(\bar{x}) \le f(\bar{x} + t(x - \bar{x})) = f(\bar{x}) + t\langle \nabla f(\bar{x}), x - \bar{x}\rangle + o(t),$$

where the inequality follows from $\bar{x}$'s being a local minimizer of f, and the equality follows from Taylor's formula. Thus, we have

$$t\langle \nabla f(\bar{x}), x - \bar{x}\rangle + o(t) \ge 0,$$

for all t small enough. For $t > 0$, dividing both sides by t and letting t go to 0 gives $\langle \nabla f(\bar{x}), x - \bar{x}\rangle \ge 0$. For $t < 0$, the same procedure leads to $\langle \nabla f(\bar{x}), x - \bar{x}\rangle \le 0$. Since an arbitrary point in L can be represented as $x - \bar{x}$ for some $x \in A$, we obtain equation (8.1).

If f is convex and x is any point in A, we have $f(x) \ge f(\bar{x}) + \langle \nabla f(\bar{x}), x - \bar{x}\rangle = f(\bar{x})$, where the inequality follows from the convexity of f and the equality follows from (8.1). □

Next, we compute the gradient and the Hessian of the function $f(X) = \ln \det X$.

Lemma 8.2. *The gradient and the Hessian of the function $f(X) = \ln \det X$ at a symmetric, positive definite matrix X are given by*

$$\nabla f(X) = X^{-1}, \quad \nabla^2 f(X) = -X^{-1} \otimes X^{-1}. \tag{8.2}$$

Proof. We expand the Taylor series of the function $f(X)$ in a given direction $D \in \mathbb{SR}^{n\times n}$,

$$\begin{aligned}\Delta f &:= f(X + tD) - f(X) = \ln \det(X + tD) - \ln \det X \\ &= \ln \det(X^{1/2}(I + tX^{-1/2}DX^{-1/2}X^{1/2}) - \ln \det X \\ &= \ln \det(I + tX^{-1/2}DX^{-1/2}).\end{aligned}$$

Define $\widehat{D} := X^{-1/2}DX^{-1/2}$. Writing the orthogonal decomposition of $\widehat{D}$ in the form $\widehat{D} = Q\Lambda Q^T$ where Q is an $n \times n$ orthogonal matrix (whose columns are the eigenvectors of $\widehat{D}$), and Λ is an $n \times n$ diagonal matrix whose elements are the eigenvalues of $\widehat{D}$, we have

$$\begin{aligned}\Delta f &= \ln \det(I + t\Lambda) = \ln \prod_{i=1}^{n}(1 + t\lambda_i) = \sum_{i=1}^{n} \ln(1 + t\lambda_i) \\ &= t \operatorname{tr} \Lambda - \frac{t^2}{2} \operatorname{tr} \Lambda^2 + o(t^3) = t \operatorname{tr}(\widehat{D}) - \frac{t^2}{2} \operatorname{tr}(\widehat{D})^2 + o(t^3) \\ &= t\langle X^{-1}, D\rangle - \frac{t^2}{2}\langle (X^{-1}DX^{-1}), D\rangle + o(t^3),\end{aligned}$$

where we used $\ln(1 + t\alpha) = t\alpha - \frac{1}{2}(t\alpha)^2 + o(t^3)$ in the fourth equation (recall that the operator $\otimes$ is defined as $(X^{-1} \otimes X^{-1})D := X^{-1}DX^{-1}$). □

References

[1] C. G. Broyden. The convergence of an algorithm for solving sparse nonlinear systems. *Math. Comp.*, 25:285–294, 1971.

[2] R. H. Byrd and J. Nocedal. A tool for the analysis of quasi-Newton methods with application to unconstrained minimization. *SIAM J. Numer. Anal.*, 26(3):727–739, 1989.

[3] R. Courant and D. Hilbert. *Methods of mathematical physics. Vol. I.* Interscience Publishers, Inc., New York, N.Y., 1953.

[4] W. C. Davidon. Variable metric method for minimization. *SIAM J. Optim.*, 1(1):1–17, 1991.

[5] J. E. Dennis, Jr. and J. J. Moré. Quasi-Newton methods, motivation and theory. *SIAM Rev.*, 19(1):46–89, 1977.

[6] J. E. Dennis, Jr. and R. B. Schnabel. Least change secant updates for quasi-Newton methods. *SIAM Rev.*, 21(4):443–459, 1979.

[7] R. Fletcher. A new approach to variable metric methods. *Comput. J.*, 13:317–322, 1970.

[8] R. Fletcher. A new variational result for quasi-Newton formulae. *SIAM J. Optim.*, 1(1):18–21, 1991.

[9] R. Fletcher. An optimal positive definite update for sparse Hessian matrices. *SIAM J. Optim.*, 5(1):192–218, 1995.

[10] R. Fletcher and M. J. D. Powell. A rapidly convergent descent method for minimization. *Comput. J.*, 6:163–168, 1963/1964.

[11] D. Goldfarb. A family of variable-metric methods derived by variational means. *Math. Comp.*, 24:23–26, 1970.

[12] J. Greenstadt. Variations on variable-metric methods. (With discussion). *Math. Comp.*, 24:1–22, 1970.

[13] Andreas Griewank. A short proof of the Dennis-Schnabel theorem. *BIT*, 22(2):252–256, 1982.

[14] J. Nocedal. Theory of algorithms for unconstrained optimization. In *Acta numerica, 1992*, Acta Numer., pages 199–242. Cambridge Univ. Press, Cambridge, 1992.

[15] D. F. Shanno. Conditioning of quasi-Newton methods for function minimization. *Math. Comp.*, 24:647–656, 1970.

[16] Ph. L. Toint. On sparse and symmetric matrix updating subject to a linear equation. *Math. Comp.*, 31(no 140):954–961, 1977.