

Vugraphs for Appendix A.

OPTIMIZATION WITH RESPECT TO A VECTOR PARAMETER

- Problems in optimization commonly arise involving
 - real and complex vector parameters
 - complex scalar parameters
 - various constraints
- Frequently the quantity to be optimized is *not* analytic
- Need to have effective “power tools” for these problems

GRADIENT WITH RESPECT TO A REAL VECTOR PARAMETER

PROBLEM

Minimize (maximize) the quantity $Q = Q(\mathbf{a})$ with respect to the real vector

parameter $\mathbf{a} = \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_N \end{bmatrix}$

APPROACH

Set

$$\nabla_{\mathbf{a}} Q \stackrel{\text{def}}{=} \begin{bmatrix} \frac{\partial Q}{\partial a_1} \\ \frac{\partial Q}{\partial a_2} \\ \vdots \\ \frac{\partial Q}{\partial a_N} \end{bmatrix} = \mathbf{0}$$

- Will develop formal rules for computing $\nabla_{\mathbf{a}} Q$.

REAL GRADIENT EXAMPLES

1. For $Q = \mathbf{b}^T \mathbf{a} = b_1 a_1 + b_2 a_2 + \cdots + b_N a_N$

$$\nabla_{\mathbf{a}} Q = \begin{bmatrix} \frac{\partial Q}{\partial a_1} \\ \frac{\partial Q}{\partial a_2} \\ \vdots \\ \frac{\partial Q}{\partial a_N} \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_N \end{bmatrix} = \mathbf{b}$$

2. For $Q = \mathbf{a}^T \mathbf{B} \mathbf{a} = \sum_j \sum_k B_{jk} a_j a_k$

By a similar procedure: $\frac{\partial Q}{\partial a_j} = \sum_k (B_{jk} + B_{kj}) a_k$

$\Rightarrow \nabla_{\mathbf{a}} Q = (\mathbf{B} + \mathbf{B}^T) \mathbf{a}$ ($= 2\mathbf{B} \mathbf{a}$ when \mathbf{B} is symmetric)

GRADIENT WITH RESPECT TO A REAL VECTOR PARAMETER FOR SOME COMMON EXPRESSIONS

Quantity Q	$\mathbf{a}^T \mathbf{b}$	$\mathbf{b}^T \mathbf{a}$	$\mathbf{a}^T \mathbf{B} \mathbf{a}$
Gradient $\nabla_{\mathbf{a}} Q$	\mathbf{b}	\mathbf{b}	$2\mathbf{B} \mathbf{a}$

Note: \mathbf{B} is assumed to be symmetric.

GRADIENT WITH RESPECT TO A COMPLEX SCALAR QUANTITY

- Functions with dependence $Q = Q(a, a^*)$ are not analytic; therefore $\frac{\partial Q}{\partial a}$ does not exist.
- If a and a^* are considered separate variables, then partial derivatives usually exist and are given by

$$\frac{\partial Q}{\partial a} = \frac{1}{2} \left(\frac{\partial Q}{\partial a_r} - j \frac{\partial Q}{\partial a_i} \right) \quad \text{and} \quad \frac{\partial Q}{\partial a^*} = \frac{1}{2} \left(\frac{\partial Q}{\partial a_r} + j \frac{\partial Q}{\partial a_i} \right)$$

GRADIENT WITH RESPECT TO A COMPLEX SCALAR (cont'd.)

- For purposes of optimization, one sets $\frac{\partial Q}{\partial a_r} = \frac{\partial Q}{\partial a_i} = 0$
- This can be done by defining

$$\nabla_{\mathbf{a}} Q \stackrel{\text{def}}{=} \frac{1}{2} \left(\frac{\partial Q}{\partial a_r} - j \frac{\partial Q}{\partial a_i} \right) \quad \text{and} \quad \nabla_{\mathbf{a}^*} Q \stackrel{\text{def}}{=} \frac{1}{2} \left(\frac{\partial Q}{\partial a_r} + j \frac{\partial Q}{\partial a_i} \right)$$

and setting *either* $\nabla_{\mathbf{a}} Q$ *or* $\nabla_{\mathbf{a}^*} Q$ to zero.

COMPLEX GRADIENT RELATIONS (SCALAR PARAMETER)

Quantity Q	a^*b	ab	$ a ^2 = aa^*$
Gradient $\nabla_a Q$	0	b	a^*
Gradient $\nabla_{a^*} Q$	b	0	a

CHECKING FOR A MINIMUM OR MAXIMUM

1. The condition

$$\nabla_a Q = 0 \quad \text{or} \quad \nabla_{a^*} Q = 0$$

determines a stationary point.

2. For a minimum or maximum, let $\nabla_{ab}^2 Q \stackrel{\text{def}}{=} \nabla_a(\nabla_b Q)$. Then the following *two* conditions must hold:

$$\left(\nabla_{aa}^2 Q\right) \cdot \left(\nabla_{a^*a^*}^2 Q\right) - \left(\nabla_{aa^*}^2 Q\right)^2 < 0$$

and

$$\nabla_{aa^*}^2 Q \begin{cases} > 0 & \text{for a minimum} \\ < 0 & \text{for a maximum} \end{cases}$$

COMPLEX GRADIENT ILLUSTRATED (SCALAR PARAMETER)

Find the complex scalar parameter a to minimize

$$Q = (\mathbf{x} - a\mathbf{y})^{*T}(\mathbf{x} - a\mathbf{y})$$

Apply the complex gradient, using results from the table:

$$Q = \mathbf{x}^{*T}\mathbf{x} - a^*\mathbf{y}^{*T}\mathbf{x} - a\mathbf{x}^{*T}\mathbf{y} + |a|^2\mathbf{y}^{*T}\mathbf{y}$$

$$\nabla_{a^*}Q = -\mathbf{y}^{*T}\mathbf{x} + a\mathbf{y}^{*T}\mathbf{y} = 0$$

This yields the result $a = \frac{\mathbf{y}^{*T}\mathbf{x}}{\mathbf{y}^{*T}\mathbf{y}}$

Note: The gradient can also be computed without expanding as

$$\nabla_{a^*}Q = -\mathbf{y}^{*T}(\mathbf{x} - a\mathbf{y})$$

COMPLEX GRADIENT (cont'd.)

The result can be further checked for a minimum. Since

$$\nabla_{a^*} Q = -\mathbf{y}^{*T}(\mathbf{x} - a\mathbf{y}) \quad \text{and} \quad \nabla_a Q = -(\mathbf{x} - a\mathbf{y})^{*T}\mathbf{y}$$

therefore

$$\nabla_{aa^*}^2 Q = \nabla_a(\nabla_{a^*} Q) = \mathbf{y}^{*T}\mathbf{y} \quad \text{while} \quad \nabla_{aa}^2 Q = \nabla_{a^*a^*}^2 Q = 0$$

Thus the two conditions for a minimum

$$\left(\nabla_{aa}^2 Q\right) \cdot \left(\nabla_{a^*a^*}^2 Q\right) - \left(\nabla_{aa^*}^2 Q\right)^2 = 0 - (\mathbf{y}^{*T}\mathbf{y})^2 < 0$$

and

$$\nabla_{aa^*}^2 Q = \mathbf{y}^{*T}\mathbf{y} > 0$$

are satisfied.

COMPLEX GRADIENT WITH RESPECT TO A VECTOR PARAMETER

$$\nabla_{\mathbf{a}} Q = (\nabla_{\mathbf{a}^*} Q^*)^* \stackrel{\text{def}}{=} \frac{1}{2} (\nabla_{\mathbf{a}_r} Q - j \nabla_{\mathbf{a}_i} Q)$$

Quantity Q	$\mathbf{a}^{*T} \mathbf{b}$	$\mathbf{b}^{*T} \mathbf{a}$	$\mathbf{a}^{*T} \mathbf{B} \mathbf{a}$
Gradient $\nabla_{\mathbf{a}} Q$	$\mathbf{0}$	\mathbf{b}^*	$(\mathbf{B} \mathbf{a})^*$
Gradient $\nabla_{\mathbf{a}^*} Q$	\mathbf{b}	$\mathbf{0}$	$\mathbf{B} \mathbf{a}$

Note: \mathbf{B} is assumed Hermitian symmetric.

CONSTRAINED OPTIMIZATION

PROBLEM

Minimize (maximize) the quantity $Q(\mathbf{a})$ subject to a complex constraint $C(\mathbf{a}) = 0$.

APPROACH

Form the Lagrangian

$$\mathcal{L} = Q(\mathbf{a}) + \lambda C(\mathbf{a}) + \lambda^* C^*(\mathbf{a})$$

and set and set the *complex gradient* $\nabla_{\mathbf{a}}\mathcal{L}$ or $\nabla_{\mathbf{a}^*}\mathcal{L}$ to zero.

CONSTRAINED OPTIMIZATION (cont'd.)

WHY IT WORKS

Observe that

$$\begin{aligned}\mathcal{L} &= Q(\mathbf{a}) + \lambda \mathcal{C}(\mathbf{a}) + \lambda^* \mathcal{C}^*(\mathbf{a}) \\ &= Q(\mathbf{a}) + 2\operatorname{Re} \lambda \mathcal{C}(\mathbf{a}) \\ &= Q(\mathbf{a}) + 2\lambda_r \mathcal{C}_r(\mathbf{a}) - 2\lambda_i \mathcal{C}_i(\mathbf{a})\end{aligned}$$

It is equivalent to adding two *real* constraints, but the first form is more convenient for use with the complex gradient.

Note: When $\mathcal{C}(\mathbf{a})$ is *real*, the last term can be dropped and λ becomes a *real* Lagrange multiplier.

CONSTRAINED OPTIMIZATION ILLUSTRATED

Find \mathbf{a} to maximize

$$Q = \mathbf{a}^{*T} \mathbf{B} \mathbf{a} \quad (\mathbf{B} \text{ Hermitian symmetric})$$

subject to the constraint $\mathbf{a}^{*T} \mathbf{a} = 1$.

The constraint is first written as

$$\mathcal{C}(\mathbf{a}) = 1 - \mathbf{a}^{*T} \mathbf{a} = 0$$

where it can be observed that $\mathcal{C}(\mathbf{a})$ is *real*.

CONSTRAINED OPTIMIZATION ILLUSTRATED (cont'd.)

The Lagrangian is

$$\mathcal{L} = \mathbf{a}^{*T} \mathbf{B} \mathbf{a} + \lambda(1 - \mathbf{a}^{*T} \mathbf{a})$$

and the complex gradient condition follows:

$$\nabla_{\mathbf{a}^*} \mathcal{L} = \mathbf{B} \mathbf{a} - \lambda \mathbf{a} = \mathbf{0} \quad \implies \quad \mathbf{B} \mathbf{a} = \lambda \mathbf{a}$$

This shows that \mathbf{a} must be an *eigenvector* of \mathbf{B} , but since

$$Q = \mathbf{a}^{*T} (\mathbf{B} \mathbf{a}) = \mathbf{a}^{*T} (\lambda \mathbf{a}) = \lambda$$

the desired eigenvector to maximize Q is the one corresponding to the *largest* eigenvalue.