Classroom Figures for the Conjugate Gradient Method Without the Agonizing Pain

Jonathan Richard Shewchuk

March 7, 1994

CMU-CS-94-125$\frac{1}{2}$

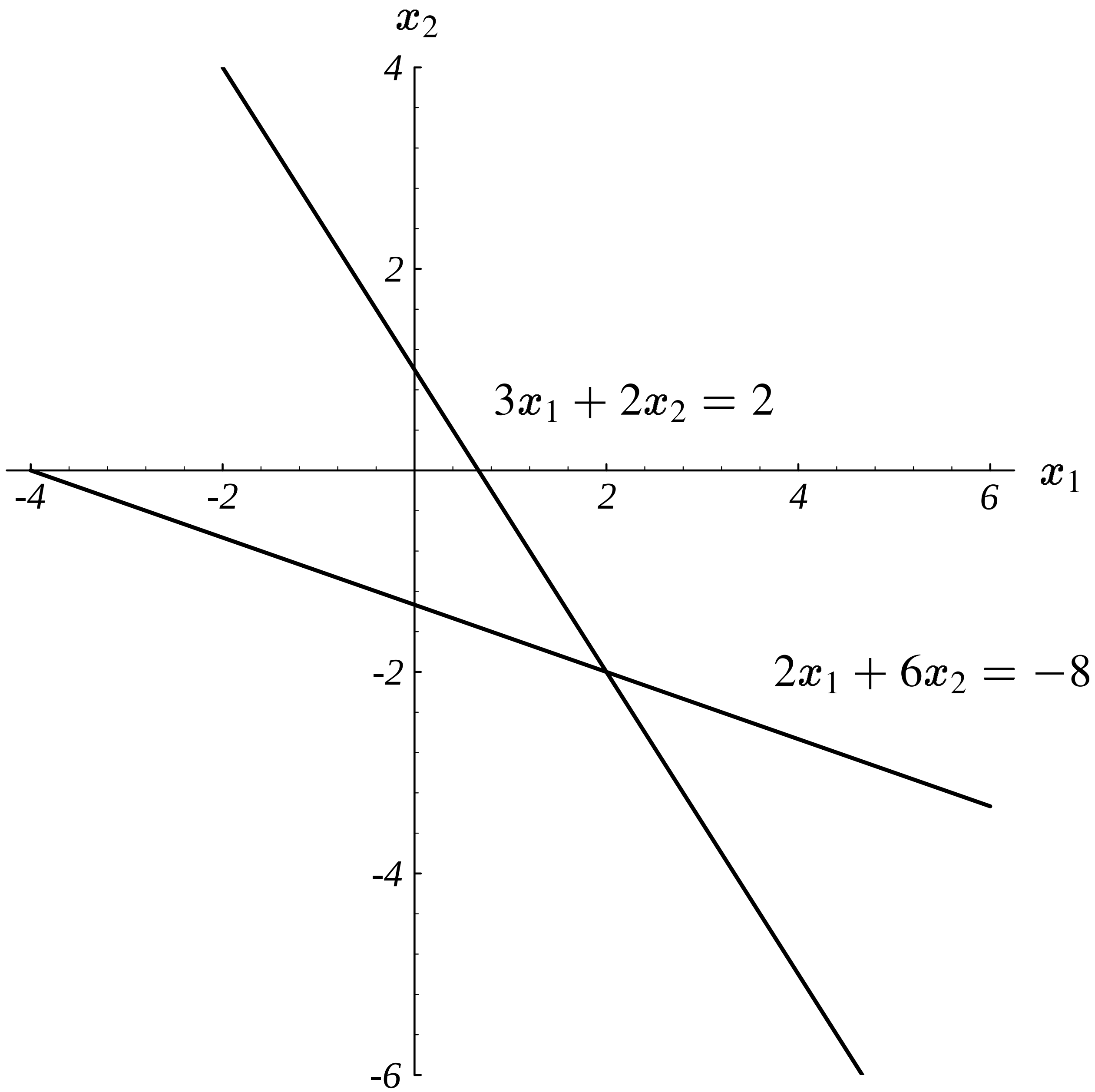
School of Computer Science
Carnegie Mellon University
Pittsburgh, PA 15213

Abstract

This report contains a set of full-page figures designed to be used as classroom transparencies for teaching from the article "An Introduction to the Conjugate Gradient Method Without the Agonizing Pain" (Technical Report CMU-CS-94-125).

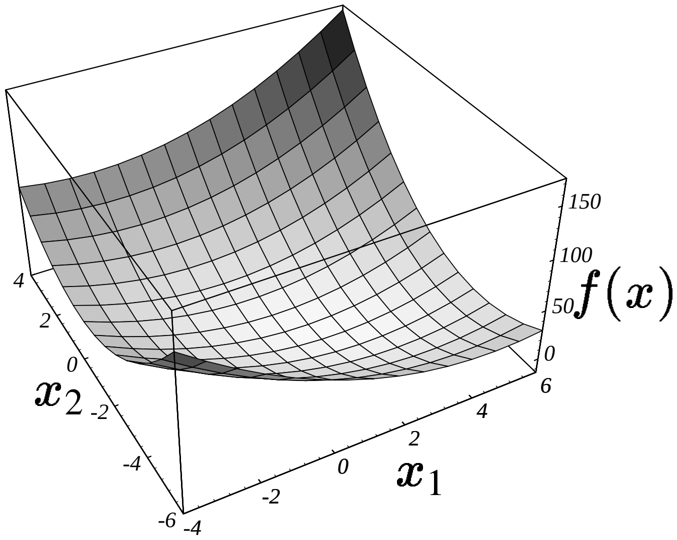
Supported in part by the Natural Sciences and Engineering Research Council of Canada under a 1967 Science and Engineering Scholarship and by the National Science Foundation under Grant ASC-9318163. The views and conclusions contained in this document are those of the author and should not be interpreted as representing the official policies, either express or implied, of NSERC, NSF, or the U.S. Government.

Keywords: conjugate gradient method, transparencies, agonizing pain

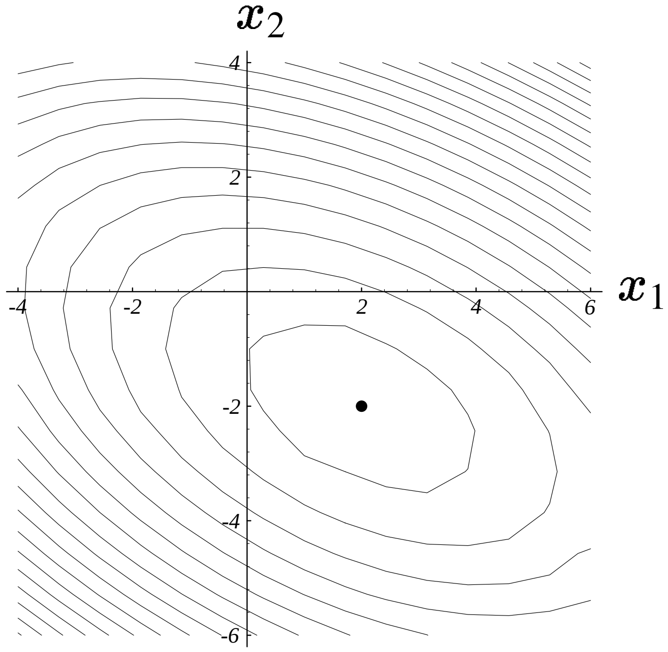


Sample 2-d linear system of the form $Ax = b$:

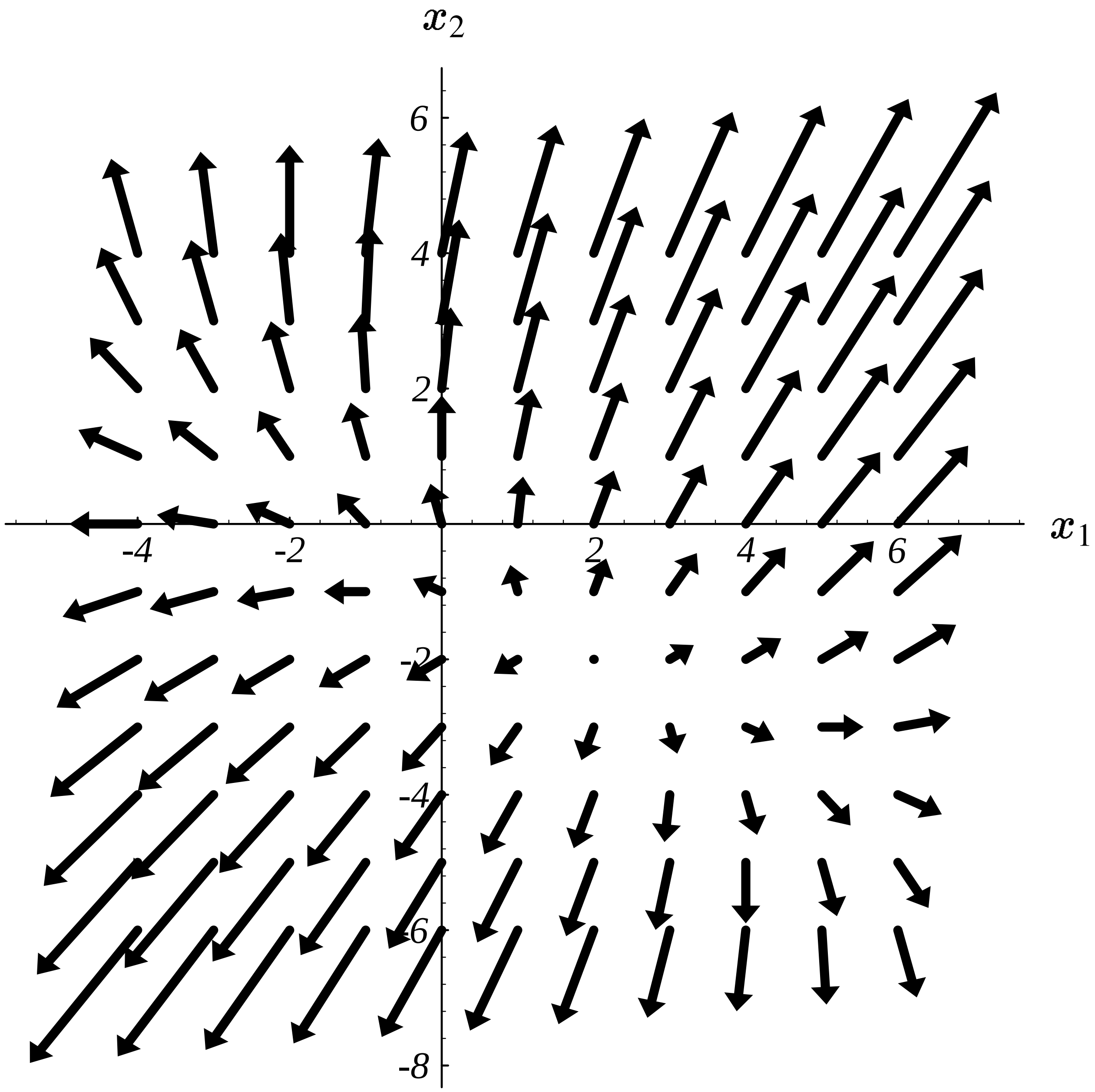
$$\begin{bmatrix} 3 & 2 \\ 2 & 6 \end{bmatrix} x = \begin{bmatrix} 2 \\ -8 \end{bmatrix}.$$



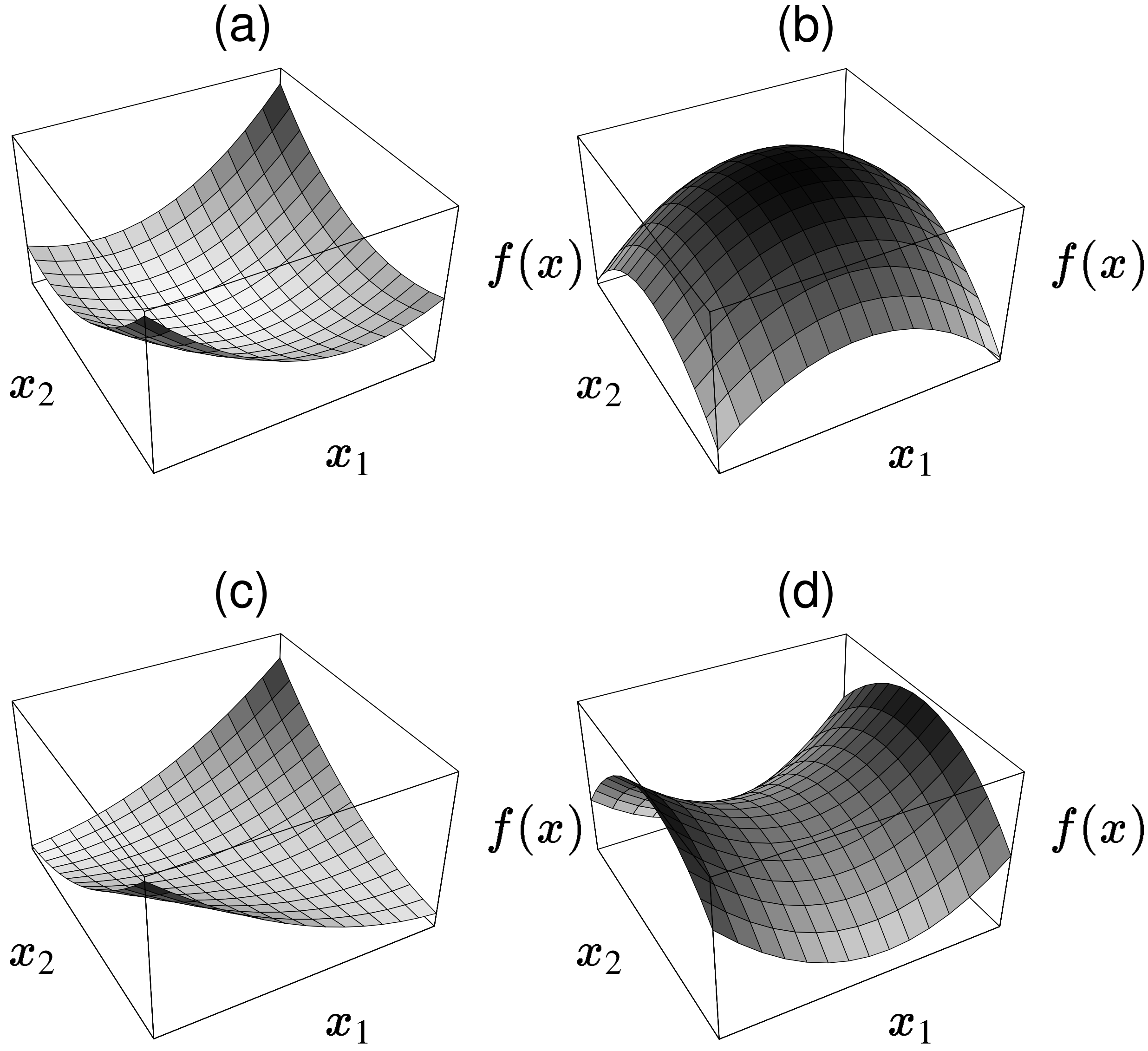
Graph of quadratic form $f(x) = \frac{1}{2}x^T Ax - b^T x + c$. The minimum point of this surface is the solution to $Ax = b$.



Contours of the quadratic form. Each ellipsoidal curve has constant $f(x)$.



Gradient $f'(x)$ of the quadratic form. For every x, the gradient points in the direction of steepest increase of $f(x)$, and is orthogonal to the contour lines.

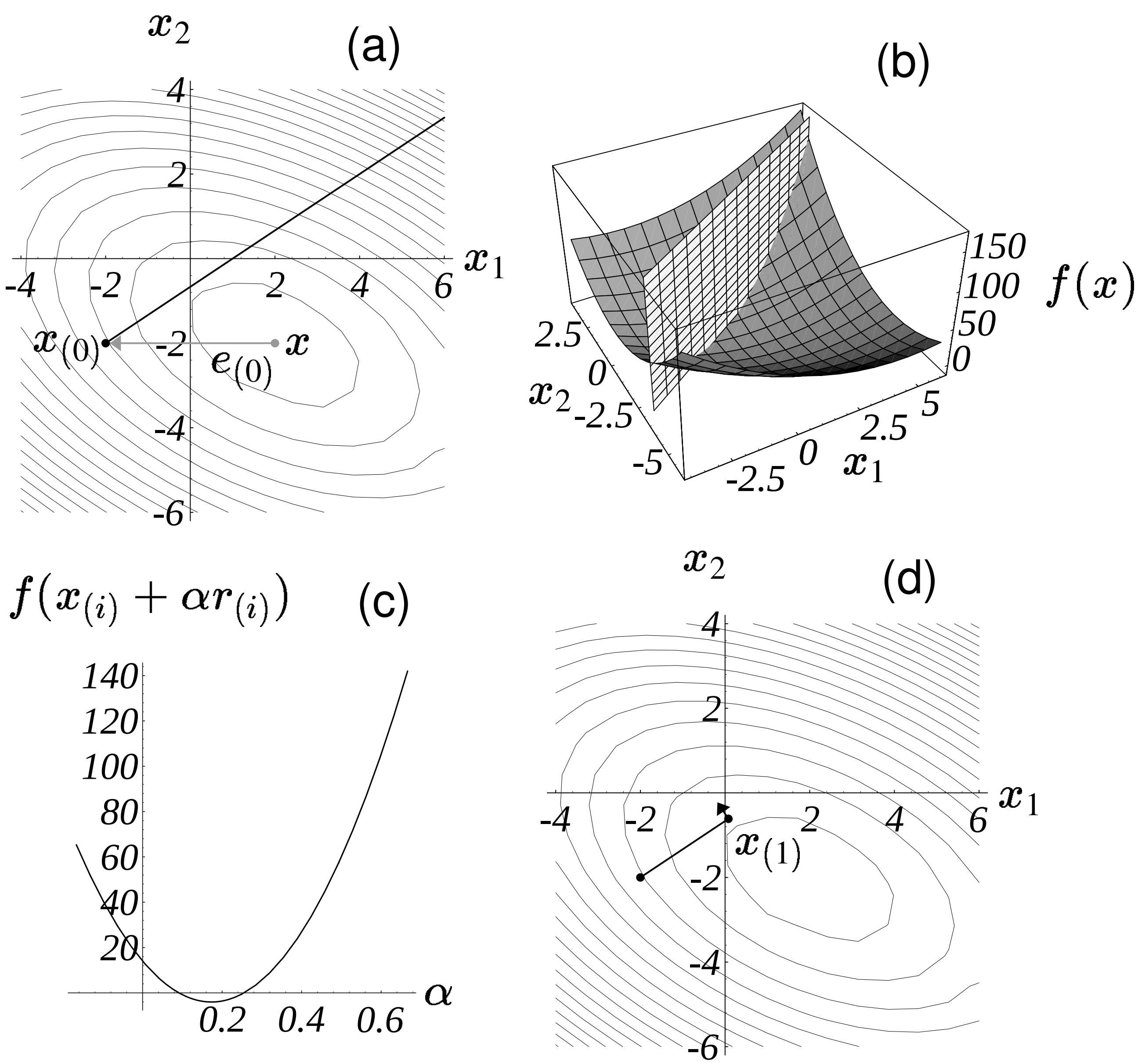


(a) Quadratic form for a positive-definite matrix.

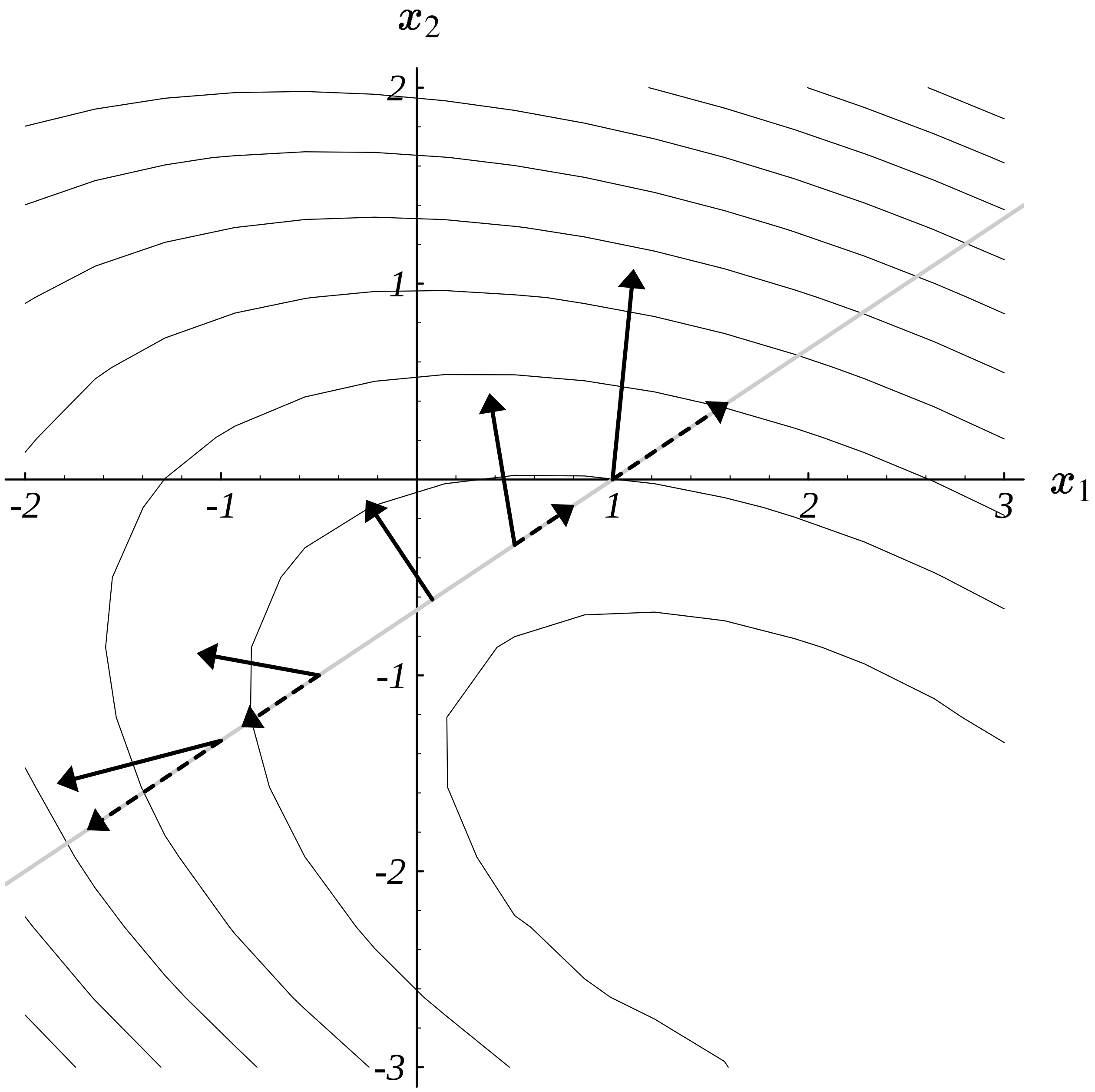
(b) For a negative-definite matrix.

(c) For a singular (and positive-indefinite) matrix. A line that runs through the bottom of the valley is the set of solutions.

(d) For an indefinite matrix.

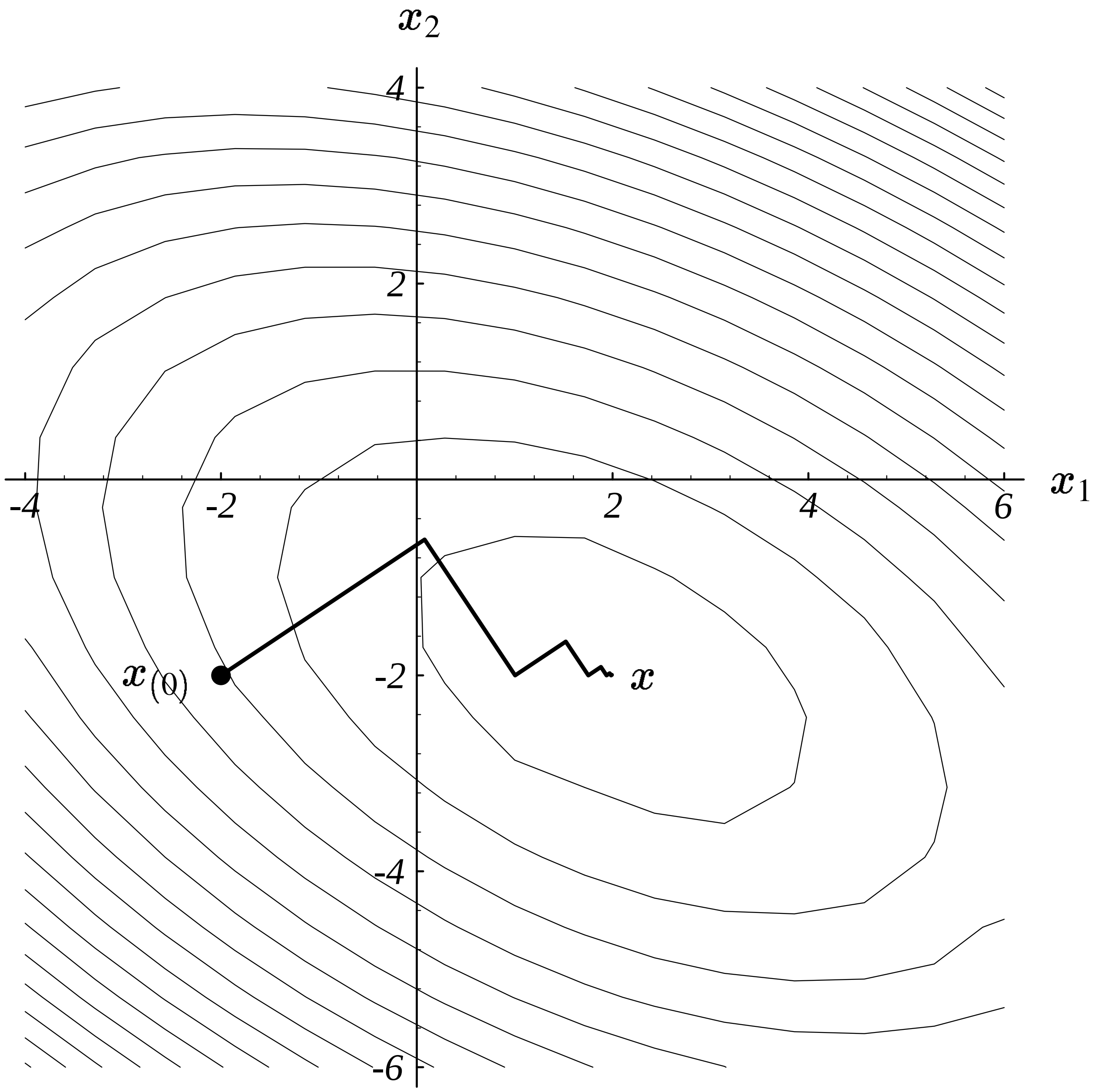


The method of Steepest Descent.

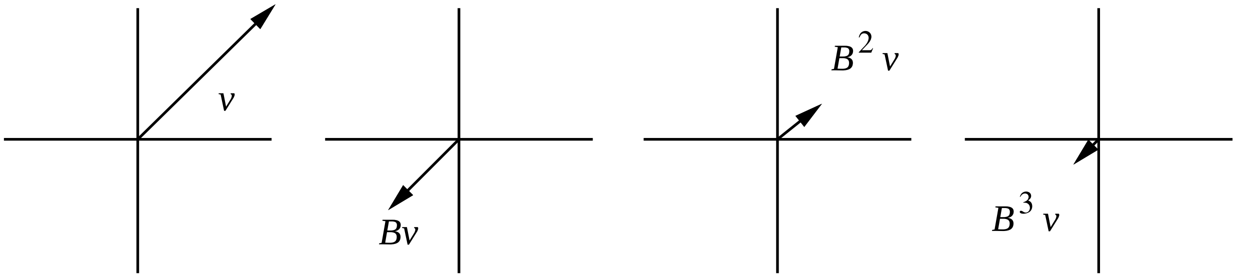


Solid arrows: Gradients.

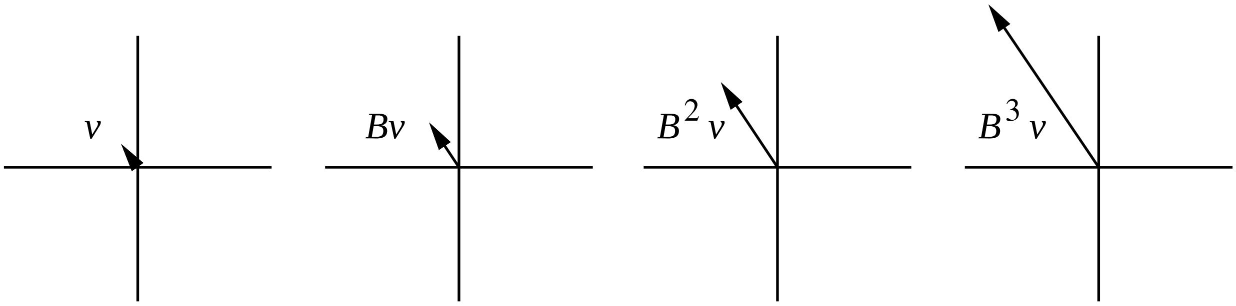
Dotted arrows: Slope along search line.



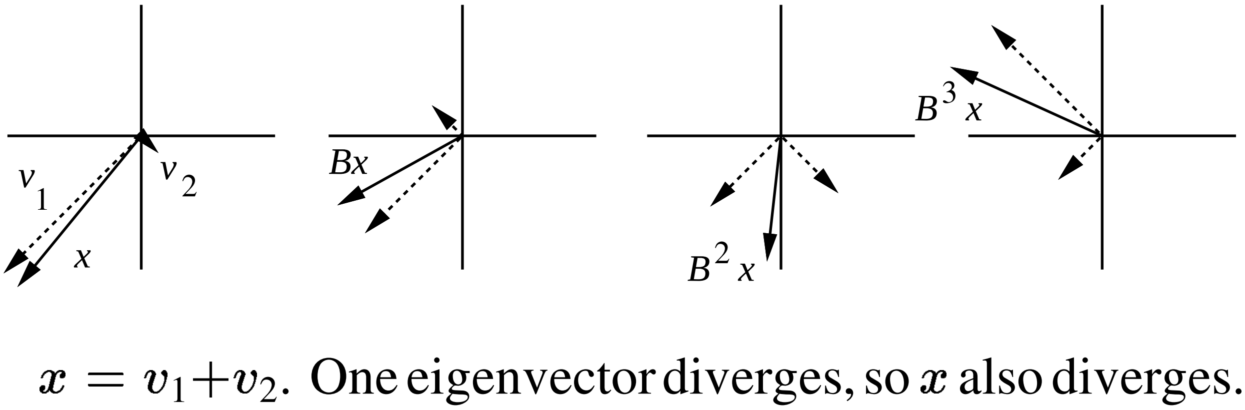
The method of Steepest Descent.



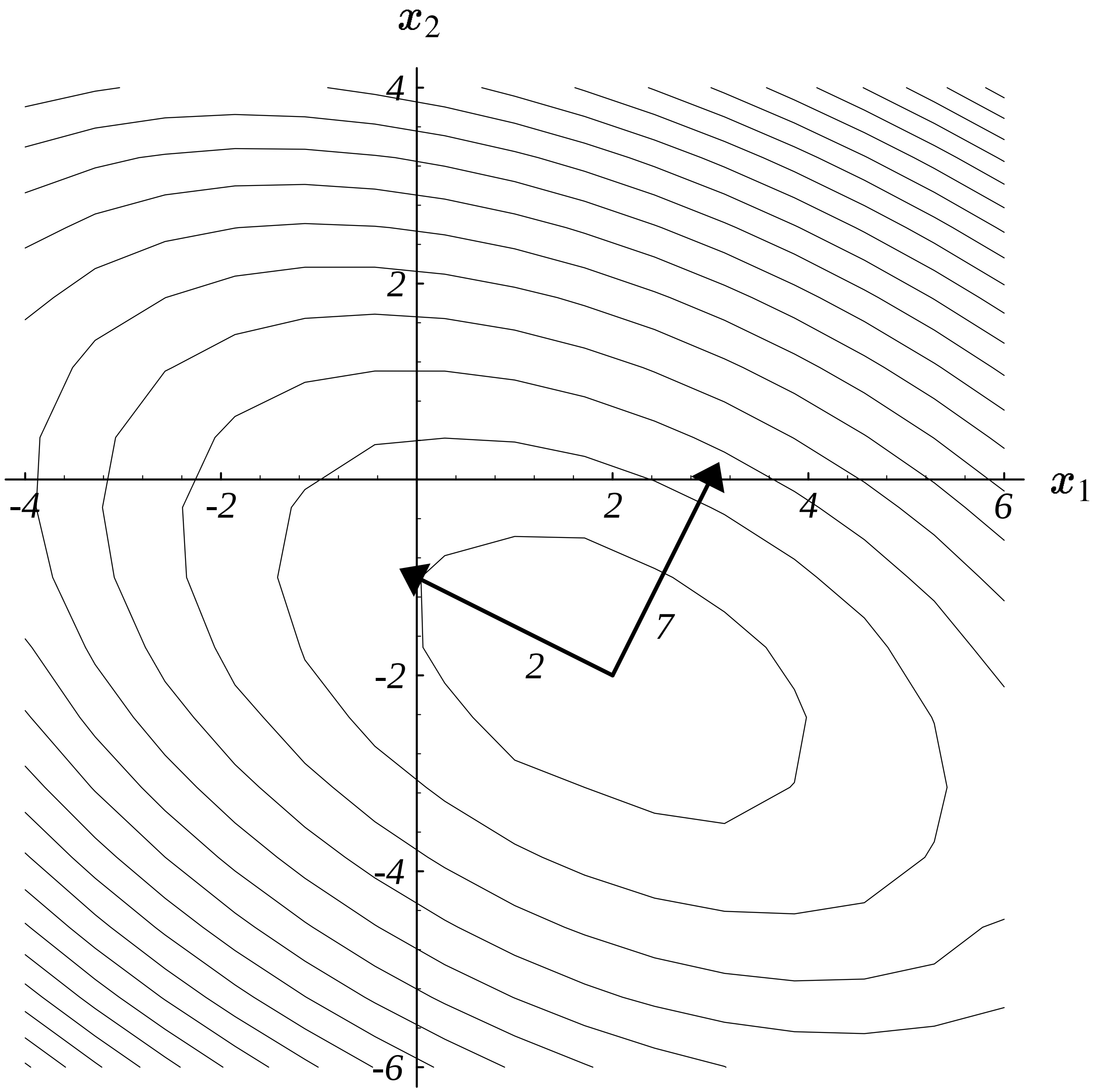
v is an eigenvector of B with a corresponding eigenvalue of -0.5. As i increases, $B^i v$ converges to zero.



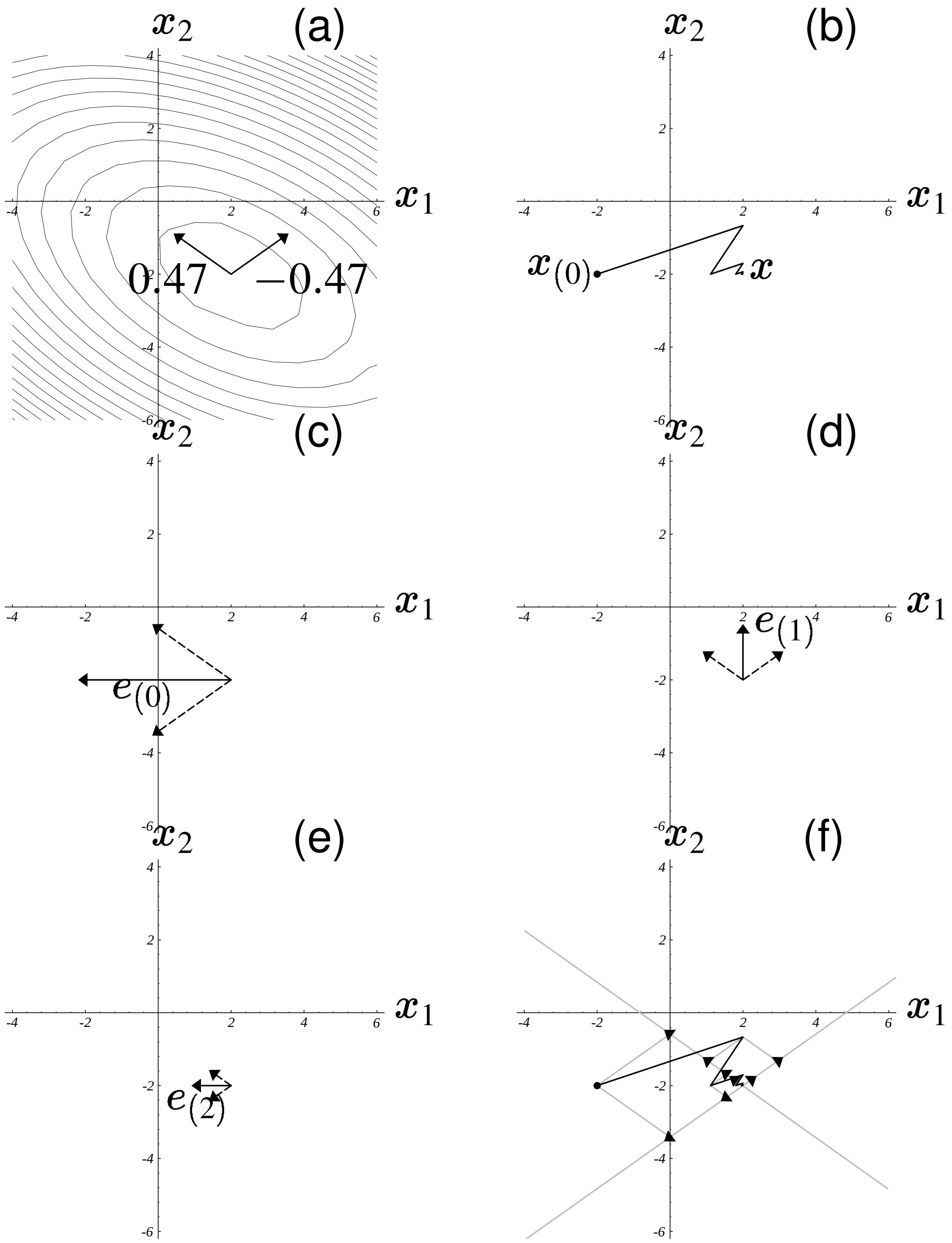
Here, v has a corresponding eigenvalue of 2. As i increases, $B^i v$ diverges to infinity.



$x = v_1 + v_2$. One eigenvector diverges, so x also diverges.

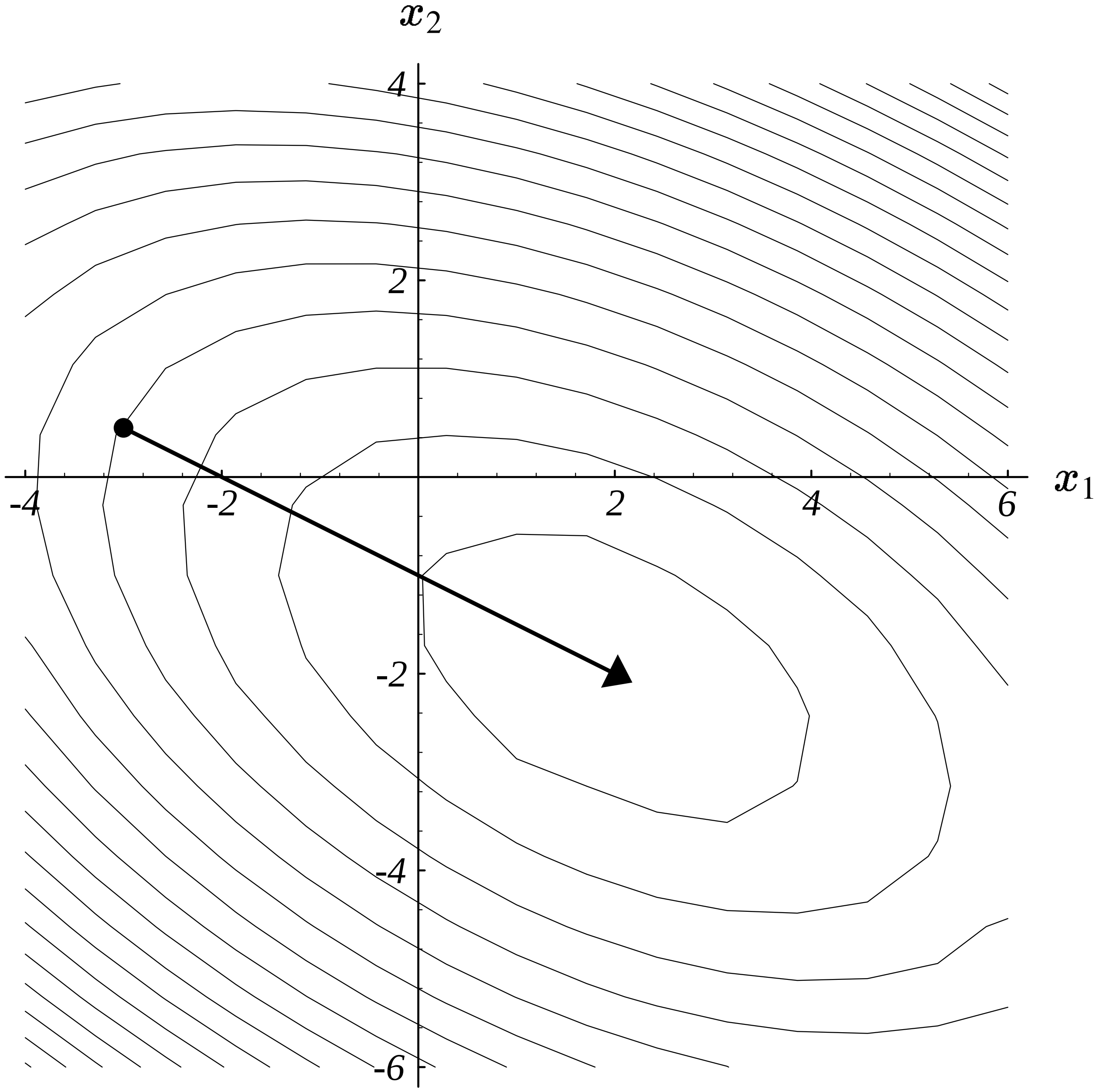


The eigenvectors of A are directed along the axes of the paraboloid defined by the quadratic form $f(x)$. Each eigenvector is labeled with its associated eigenvalue.

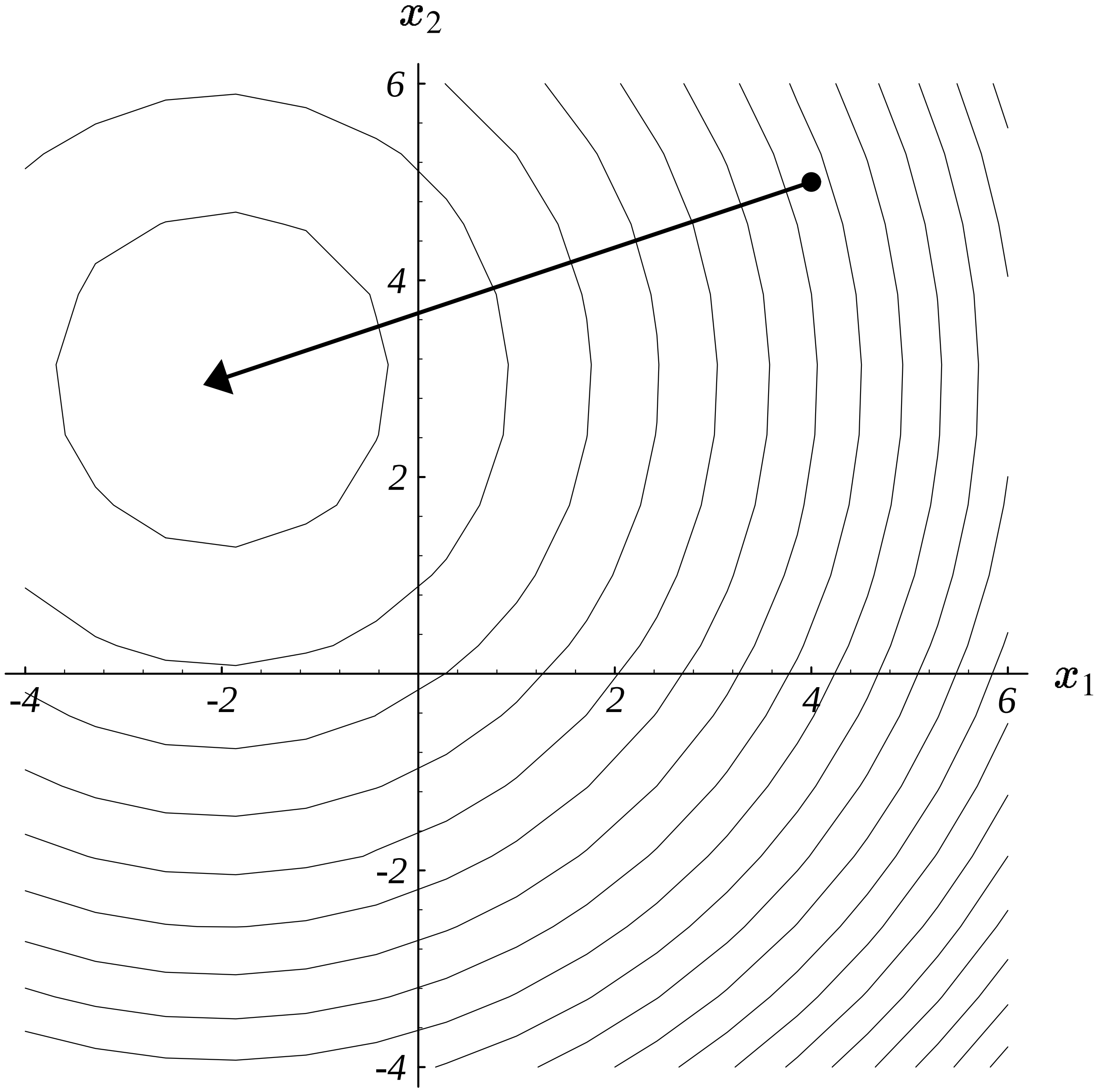


Convergence of the Jacobi Method.

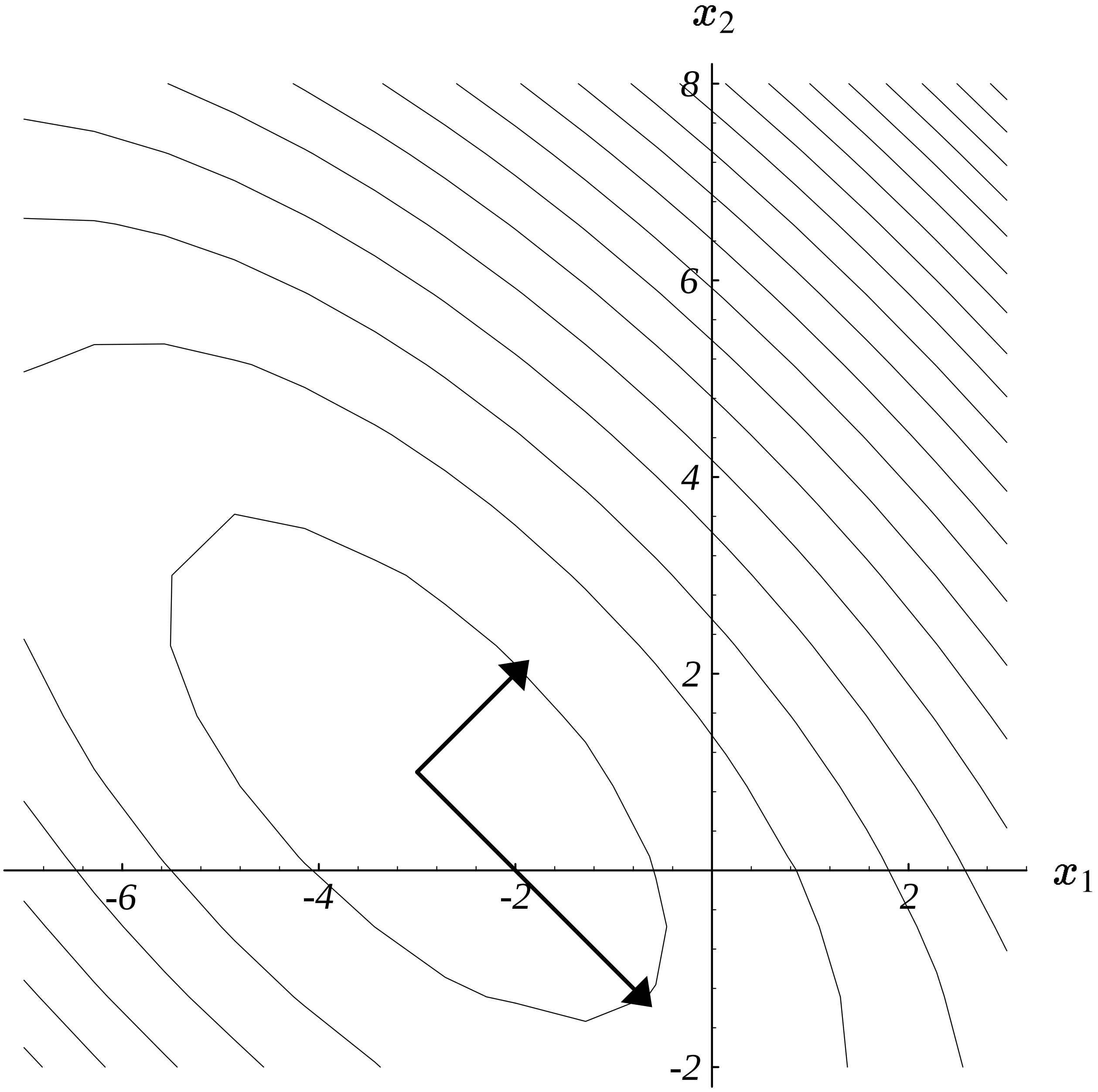
In (a), the eigenvectors of B are shown with their corresponding eigenvalues. These eigenvectors are NOT the axes of the paraboloid.



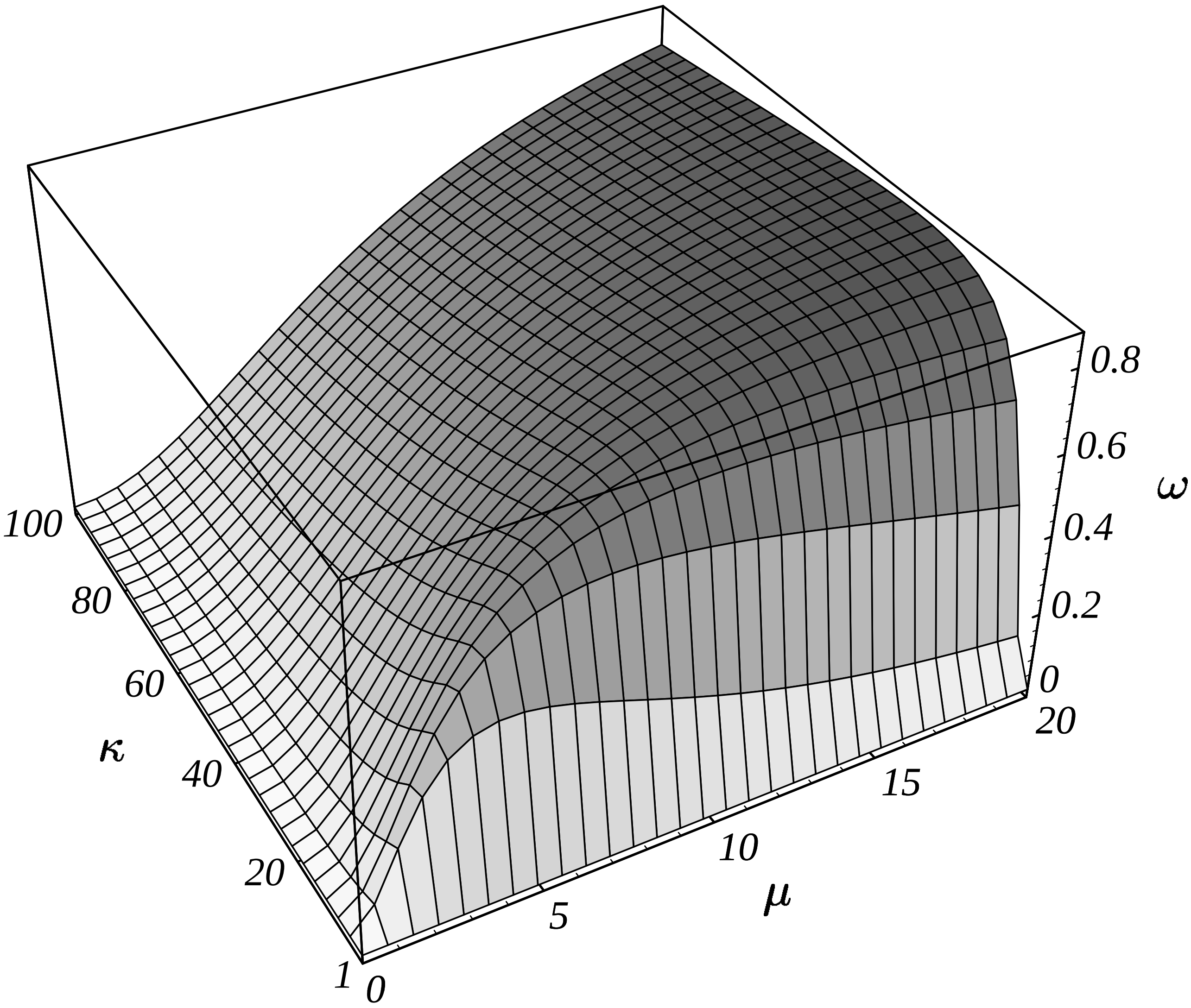
Steepest Descent converges to the exact solution on the first iteration if the error term is an eigenvector.



Steepest Descent converges to the exact solution on the first iteration if the eigenvalues are all equal.



The energy norm of these two vectors is equal.

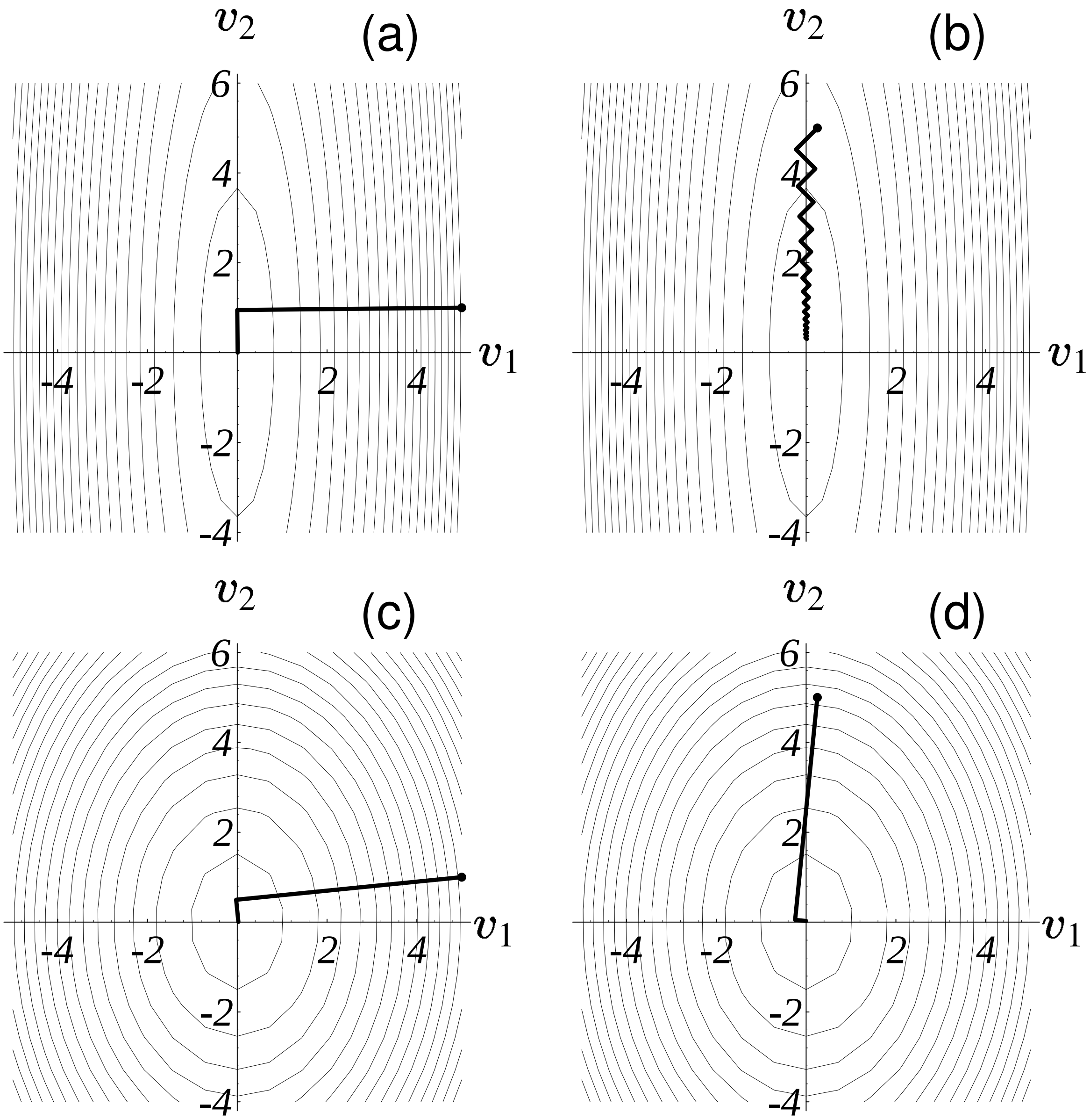


Convergence ω of Steepest Descent.

μ is the slope of $e_{(i)}$ with respect to the eigenvector axes.

κ is the condition number of A.

Convergence is worst when $\mu = \pm\kappa$.

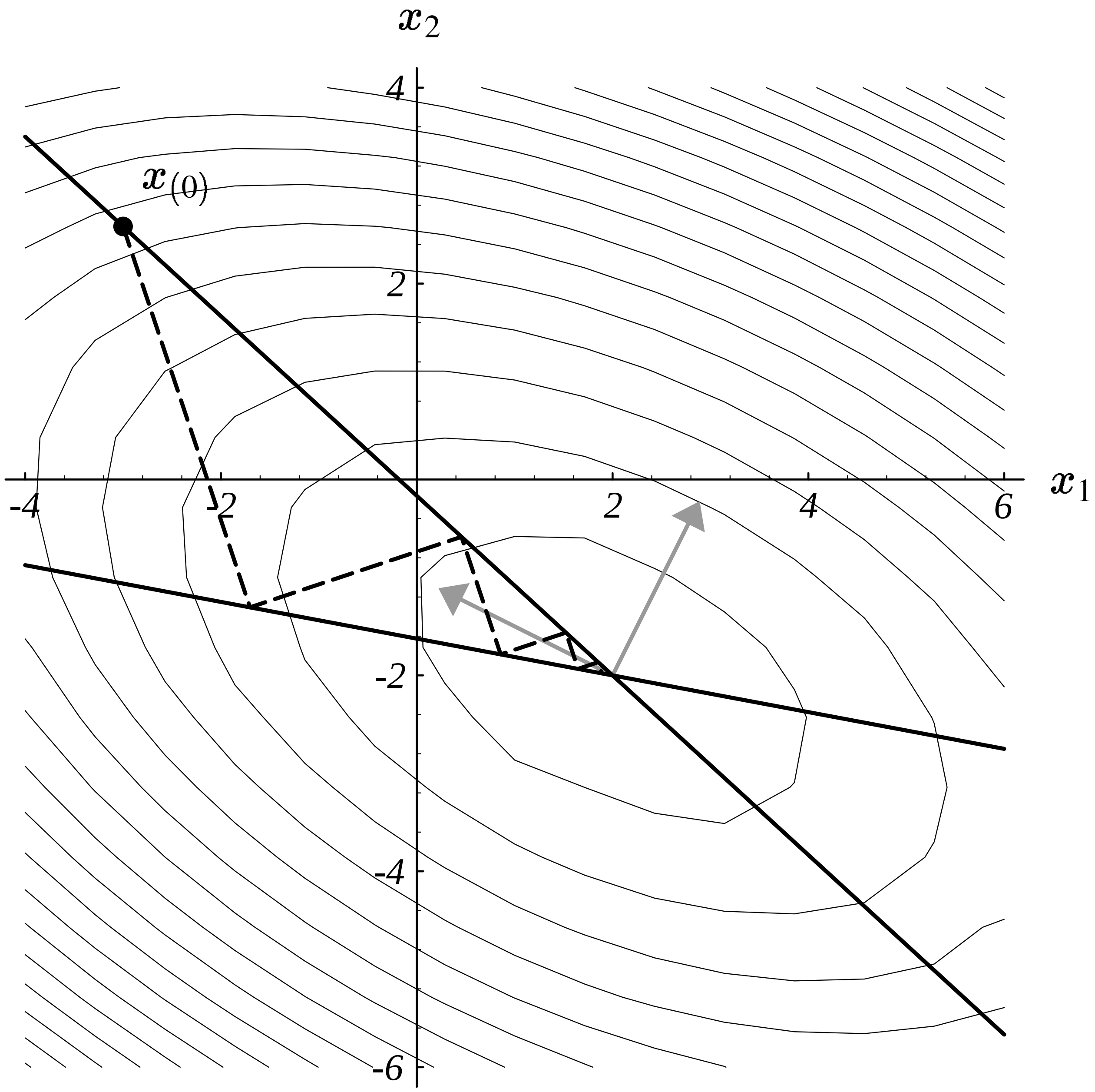


(a) Large κ, small μ.

(b) An example of poor convergence. κ and μ are both large.

(c) Small κ and μ.

(d) Small κ, large μ.

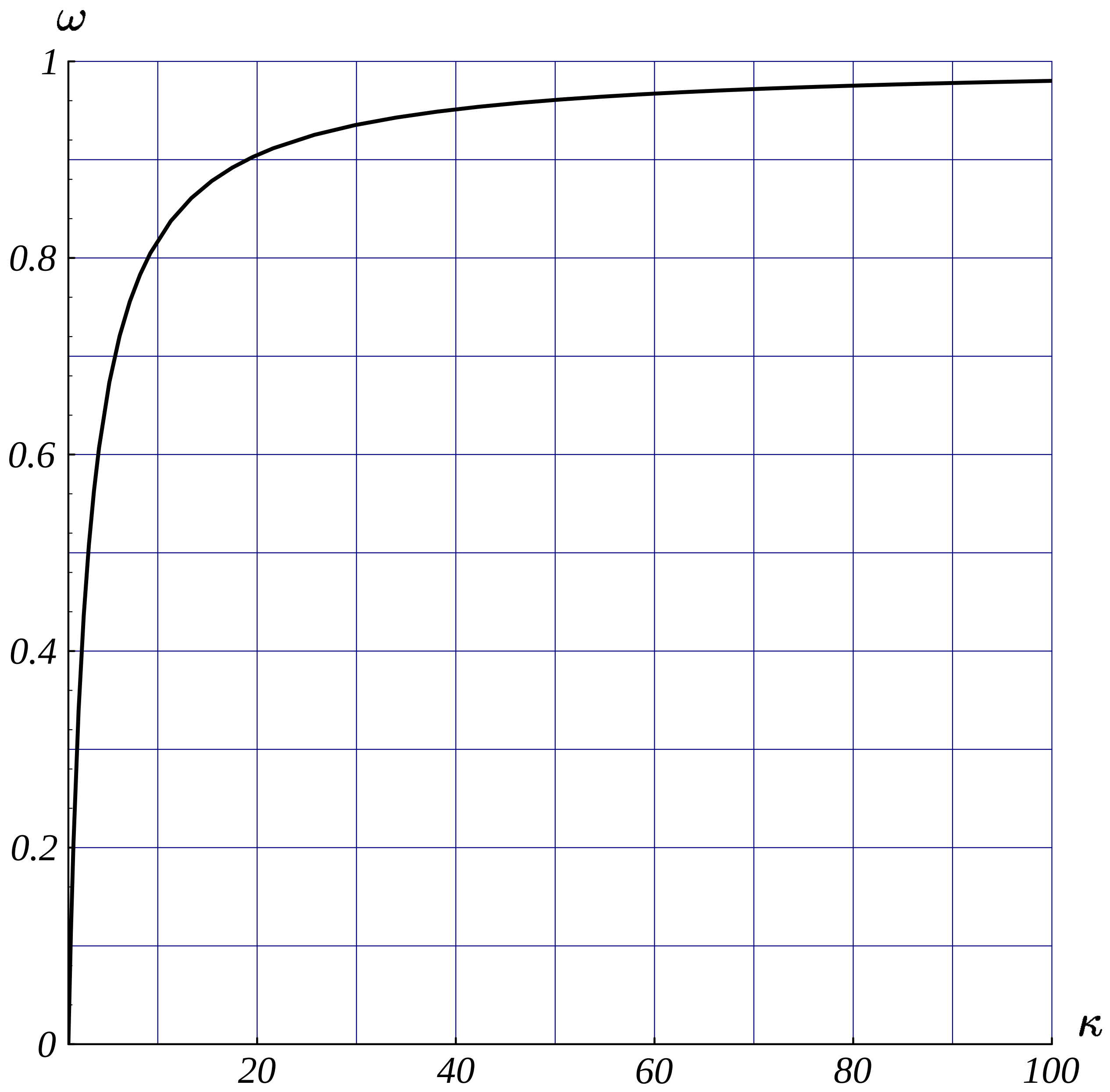


Solid lines: Worst starting points for Steepest Descent.

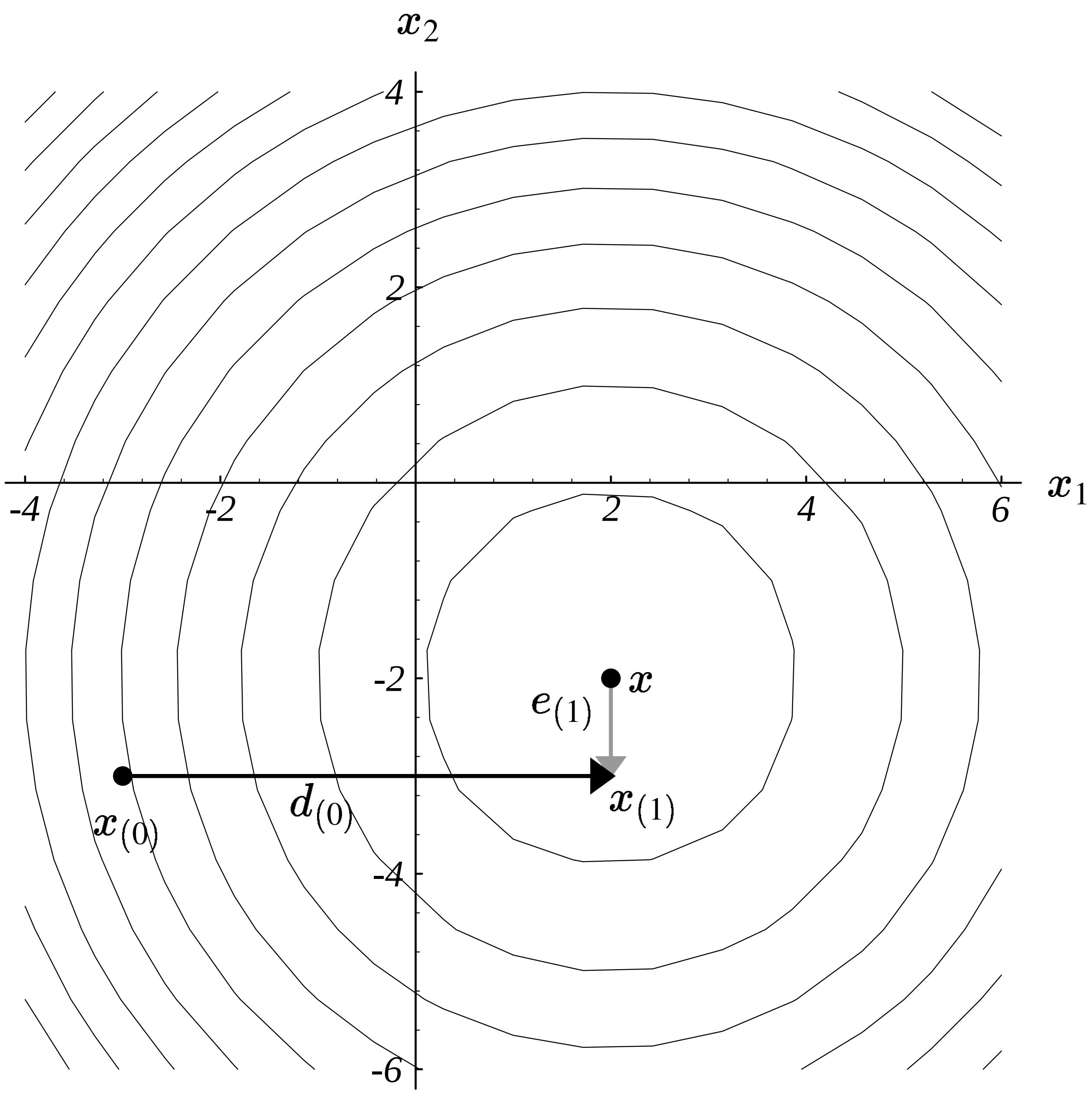
Dashed lines: Steps toward convergence.

Grey arrows: Eigenvector axes.

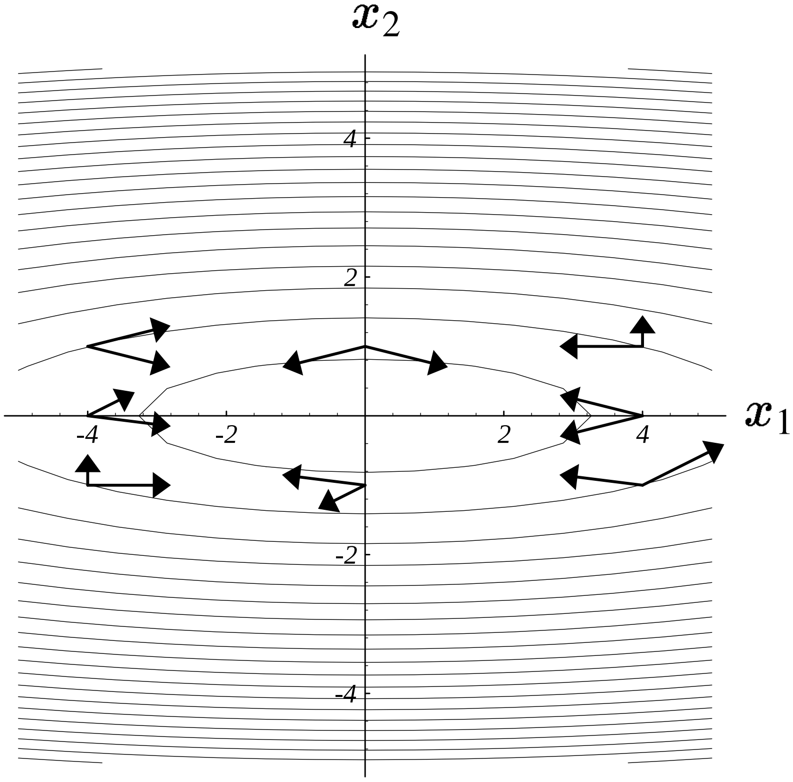
Here, $\kappa = 3.5$.



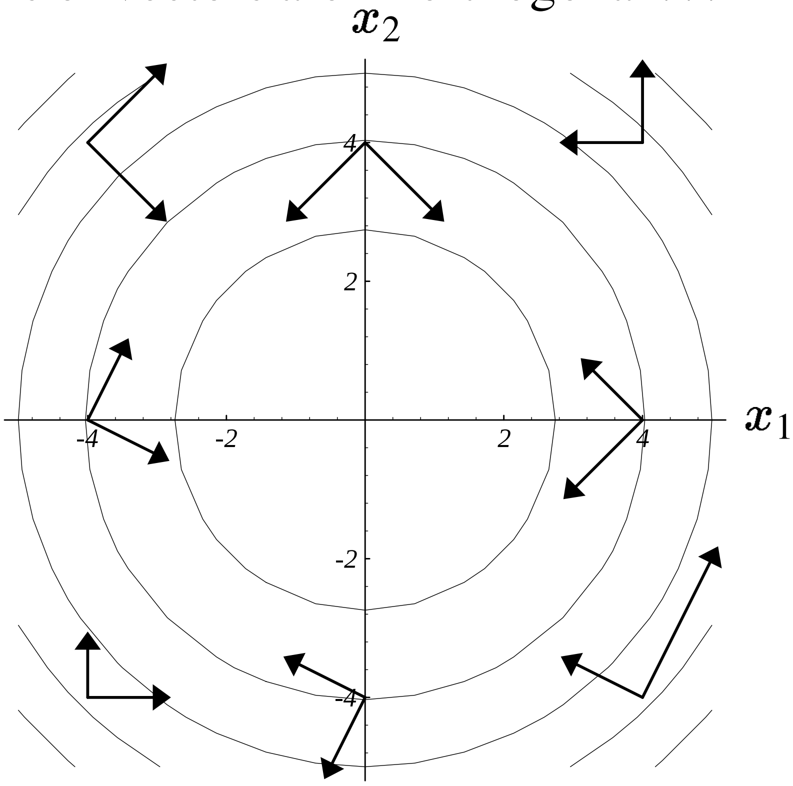
Convergence of Steepest Descent (per iteration) worsens as the condition number of the matrix increases.



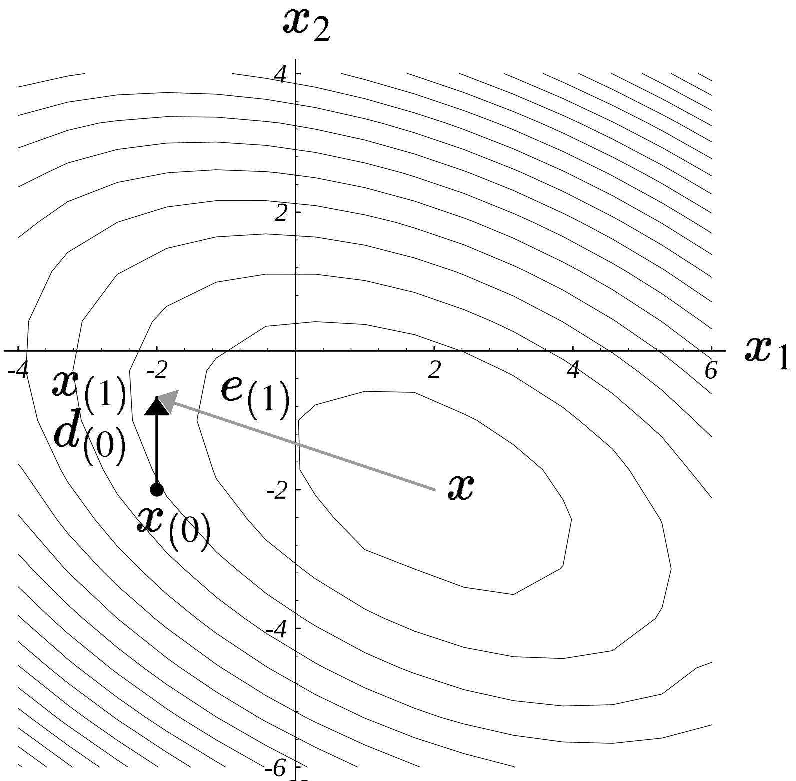
The Method of Orthogonal Directions.

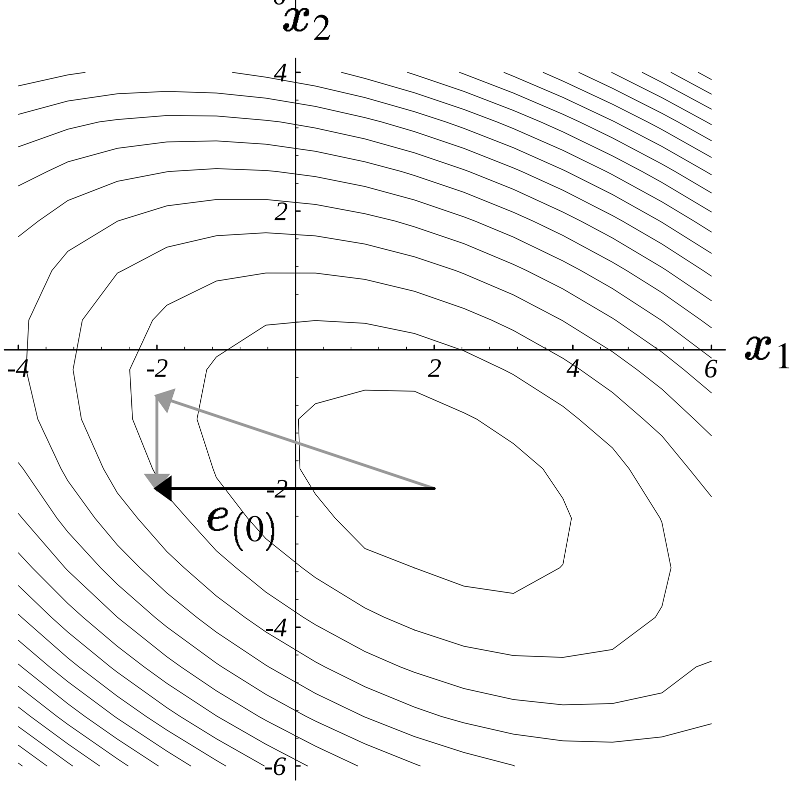


These pairs of vectors are A-orthogonal …



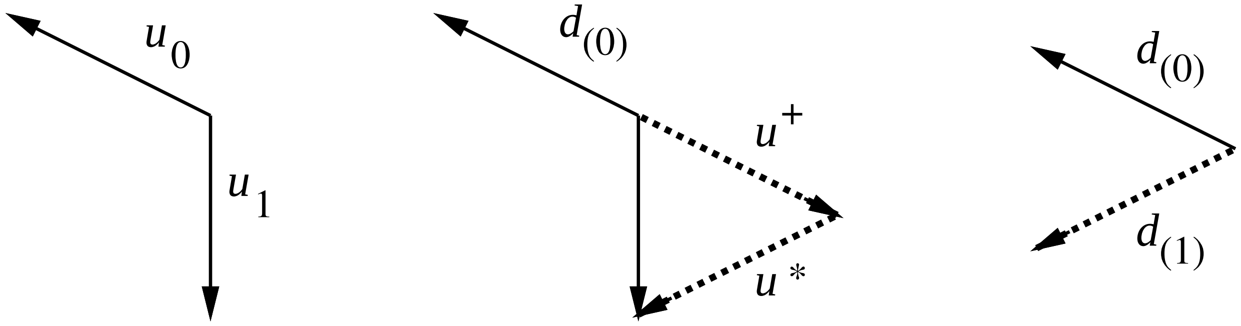
… because these pairs of vectors are orthogonal.



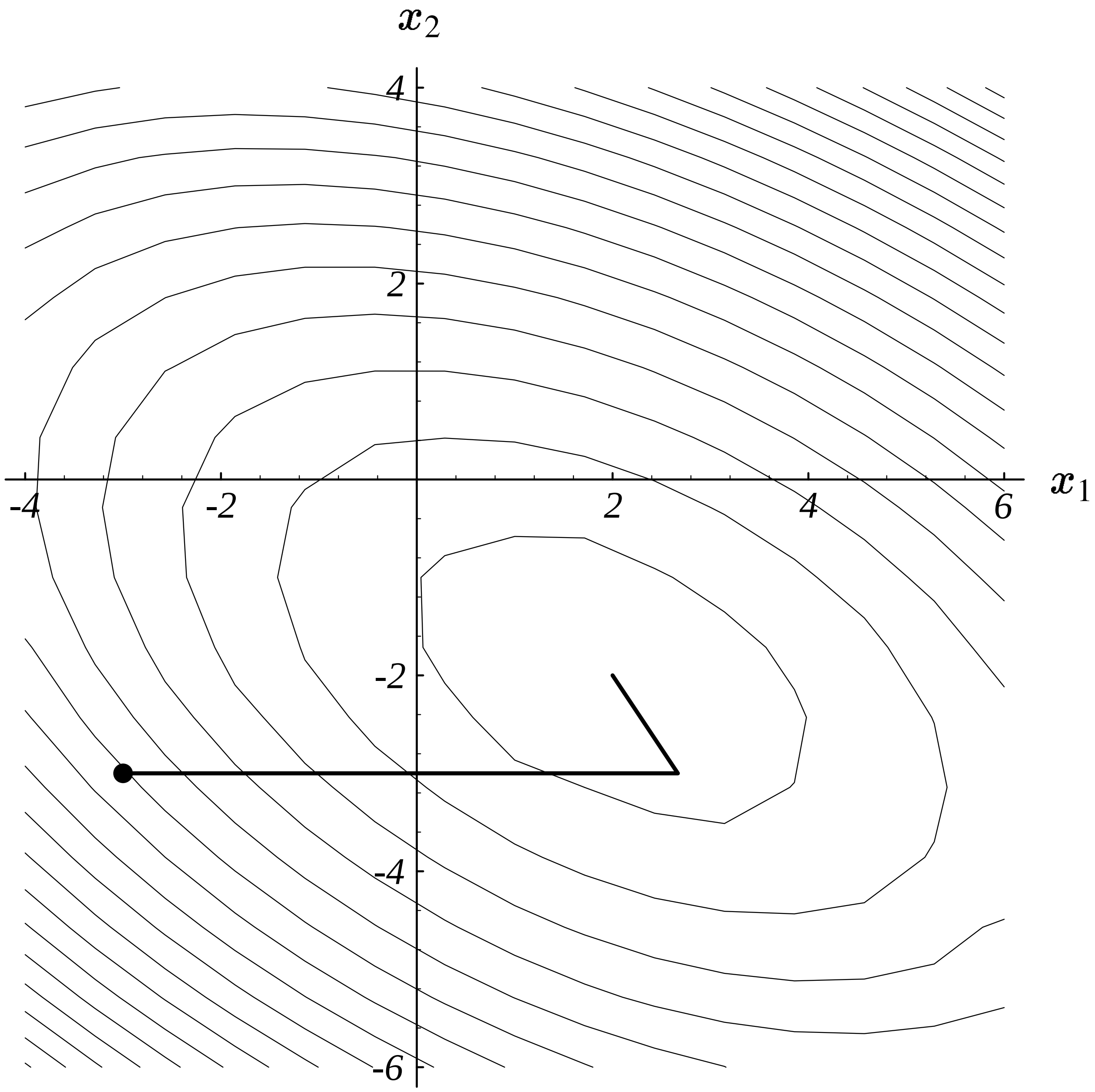


The method of Conjugate Directions converges in n steps.

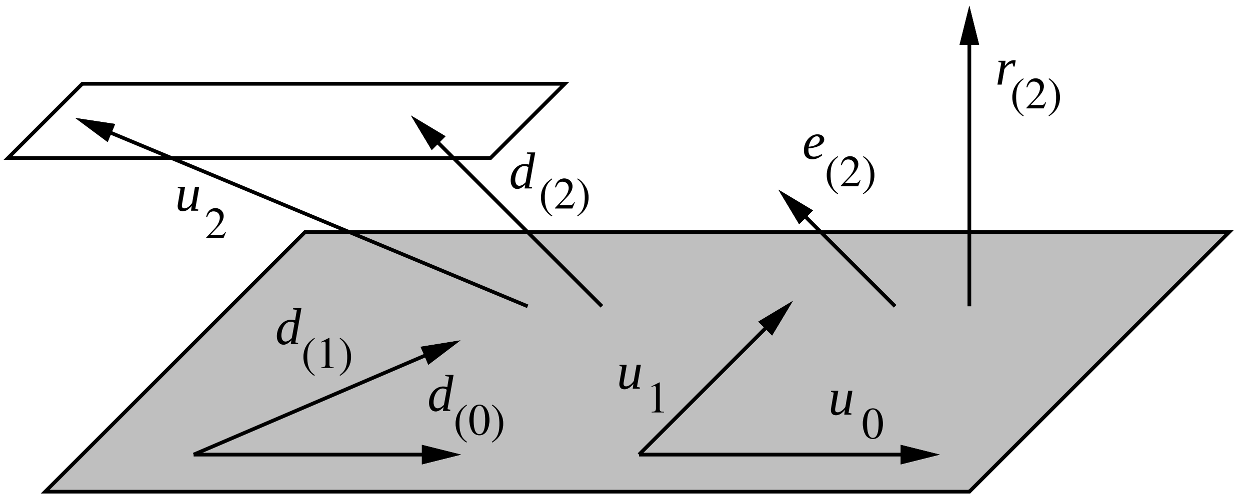
$e_{(1)}$ must be A-orthogonal to $d_{(0)}$.



Gram-Schmidt conjugation of two vectors.



The method of Conjugate Directions using the axial unit vectors, also known as Gaußian elimination.

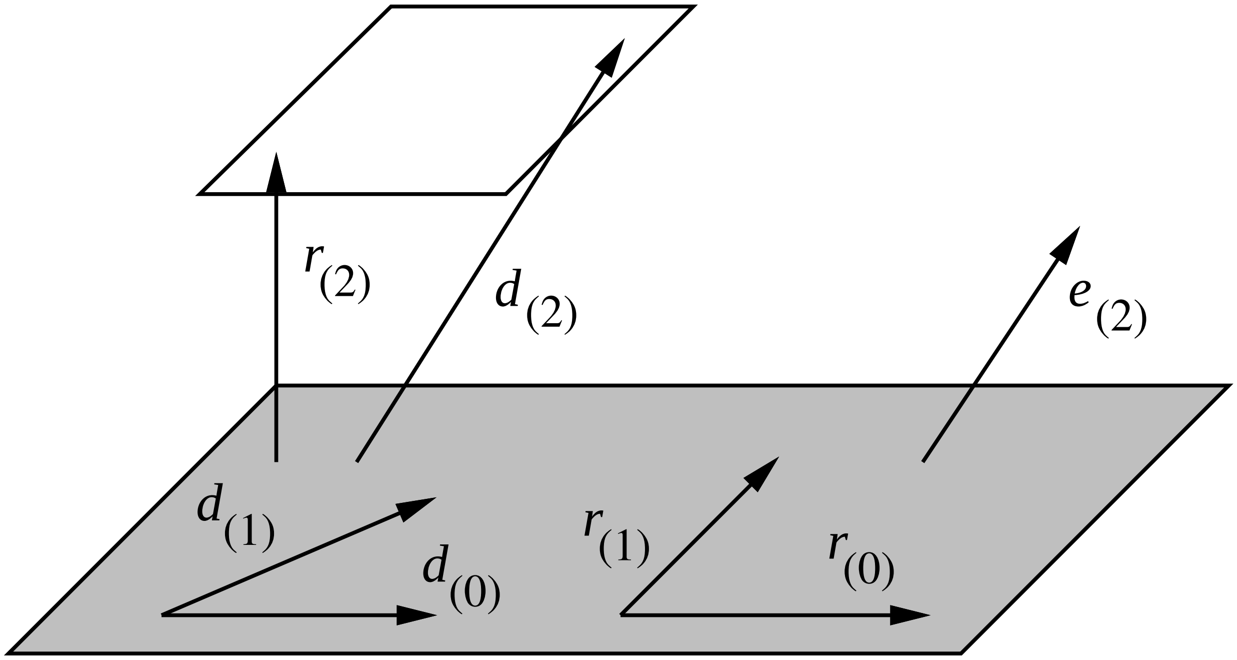


$d_{(0)}$ and $d_{(1)}$ span the same subspace as u_0, u_1 (grey-colored plane).

$e_{(2)}$ is A-orthogonal to this subspace.

$r_{(2)}$ is orthogonal to this subspace.

$d_{(2)}$ is constructed (from u_2) to be A-orthogonal to this subspace.

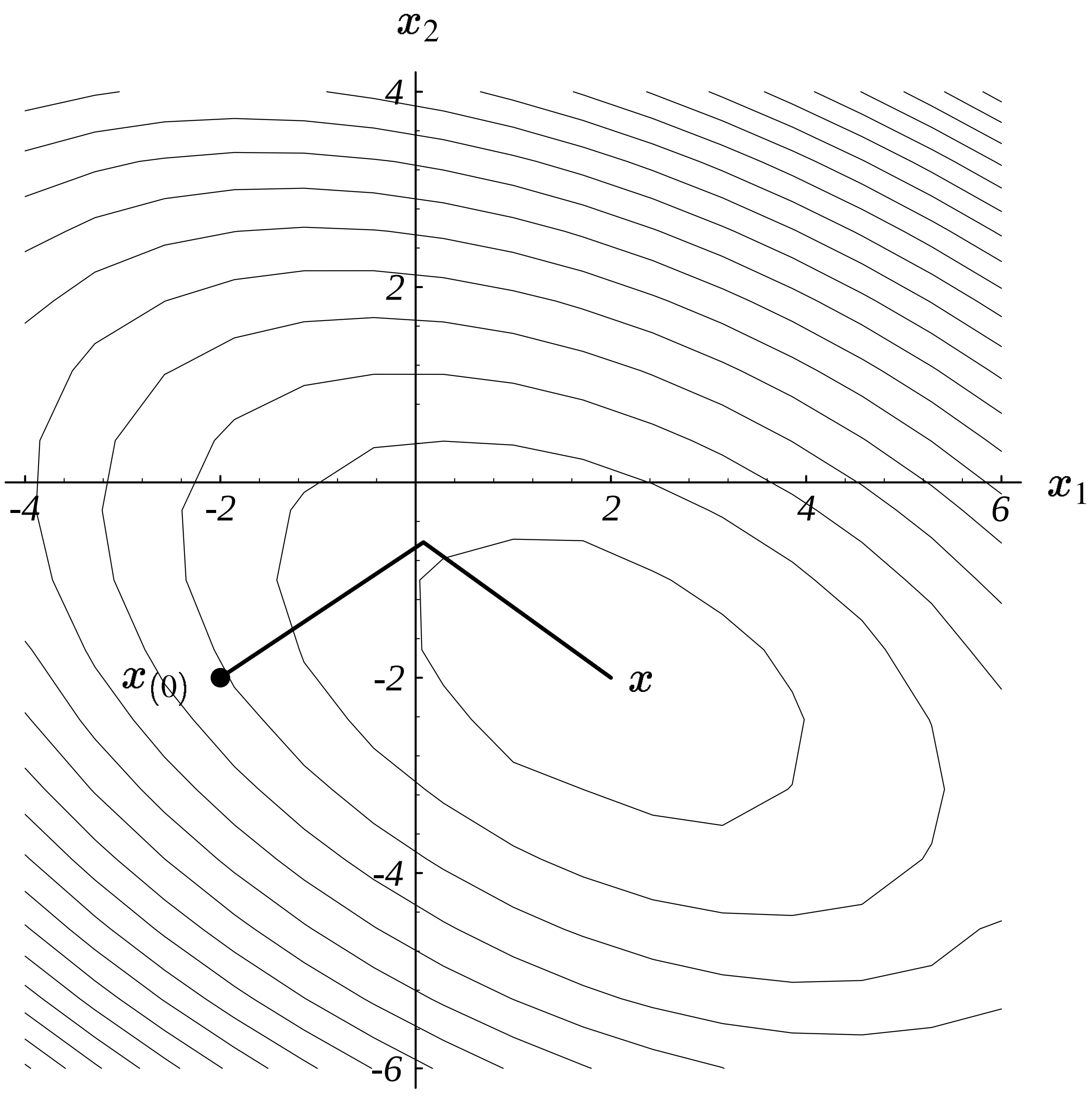


$d_{(0)}$ and $d_{(1)}$ span the same subspace as $r_{(0)}, r_{(1)}$ (grey-colored plane).

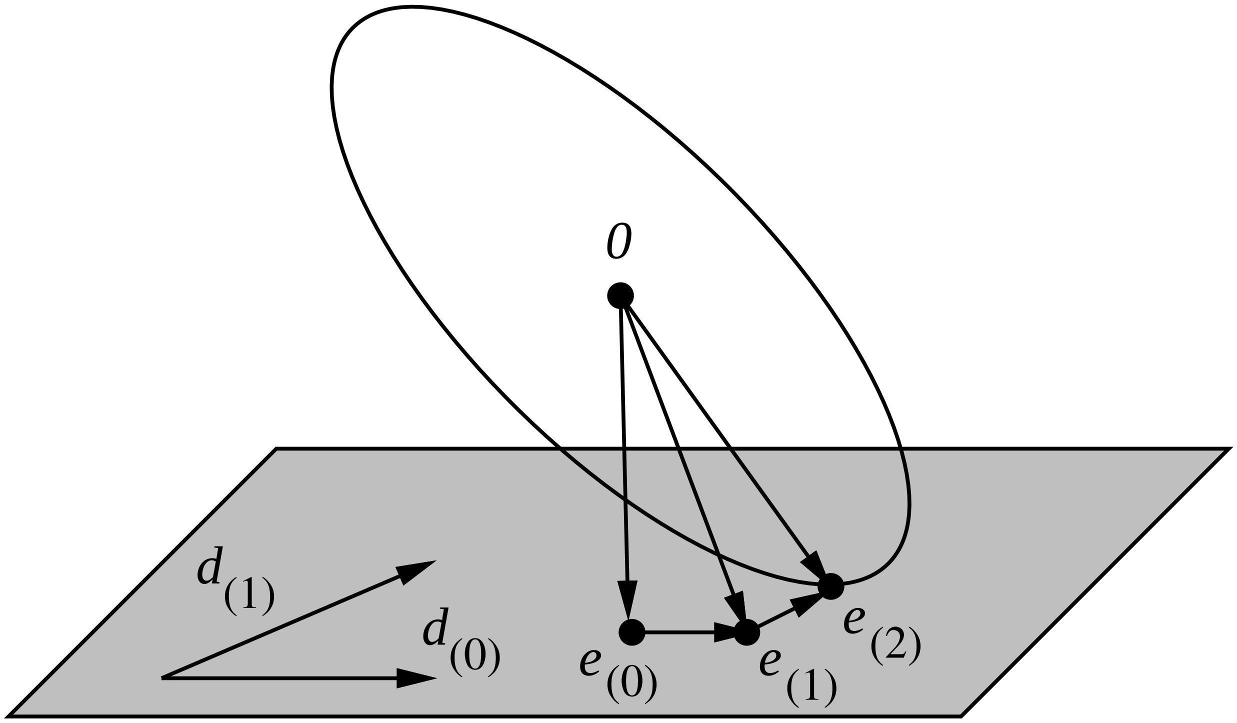
$e_{(2)}$ is A-orthogonal to this subspace.

$r_{(2)}$ is orthogonal to this subspace.

$d_{(2)}$ is constructed (from $r_{(2)}$) to be A-orthogonal to this subspace.



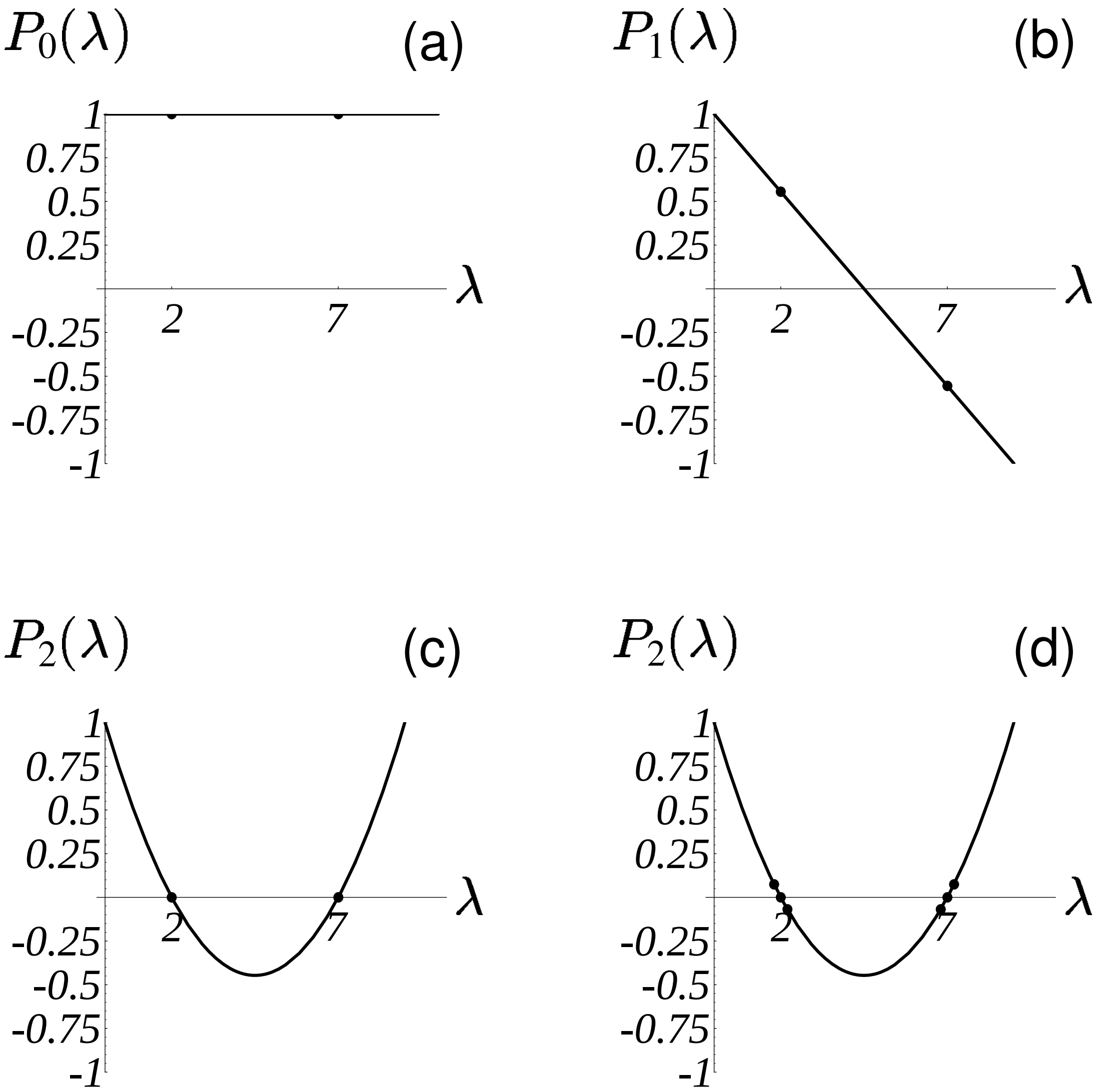
The method of Conjugate Gradients.



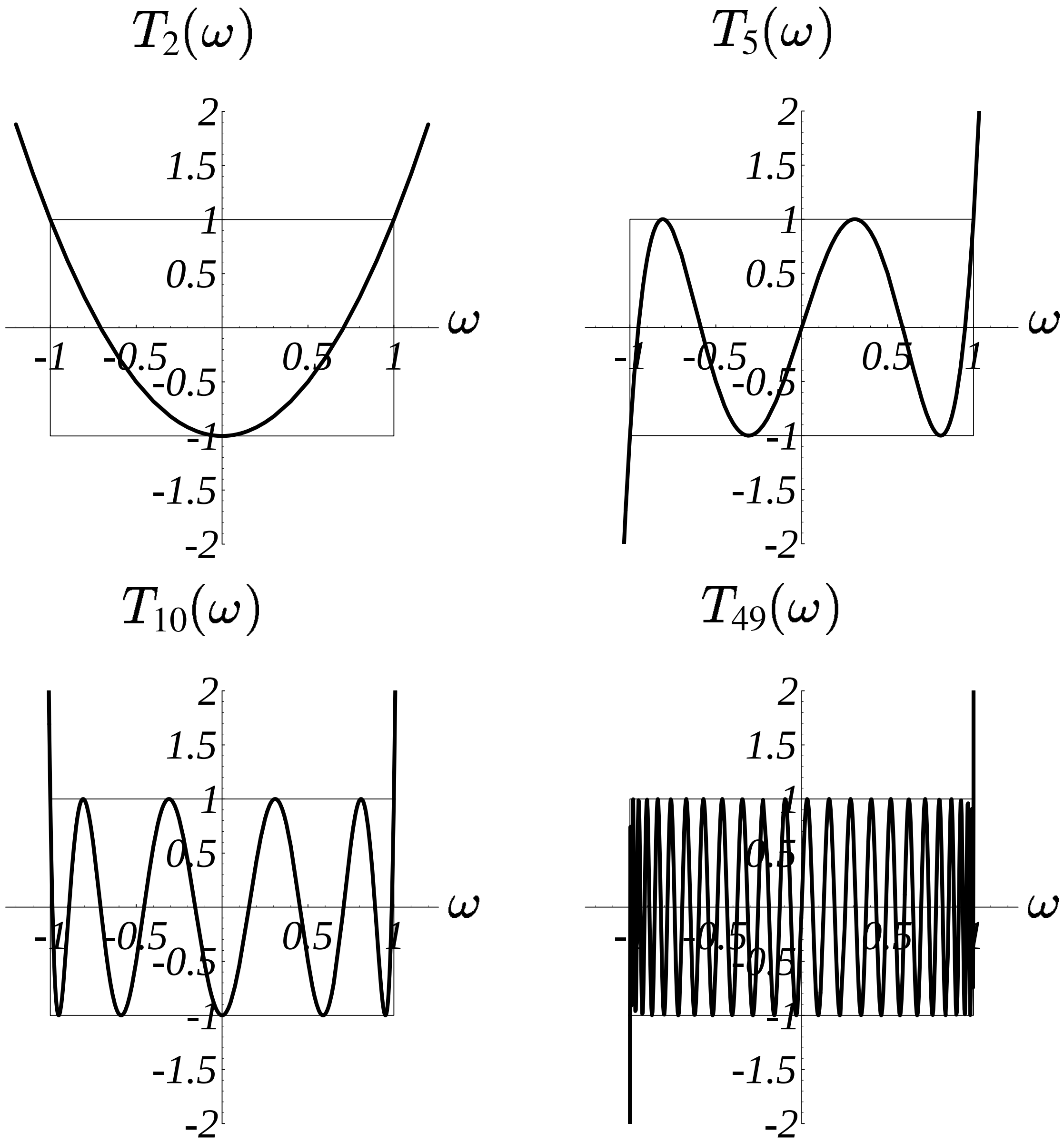
The shaded area is $\mathcal{E} = e_{(0)} + \text{span}\{d_{(0)}, d_{(1)}\}$.

The ellipsoid is a contour on which the energy norm is constant.

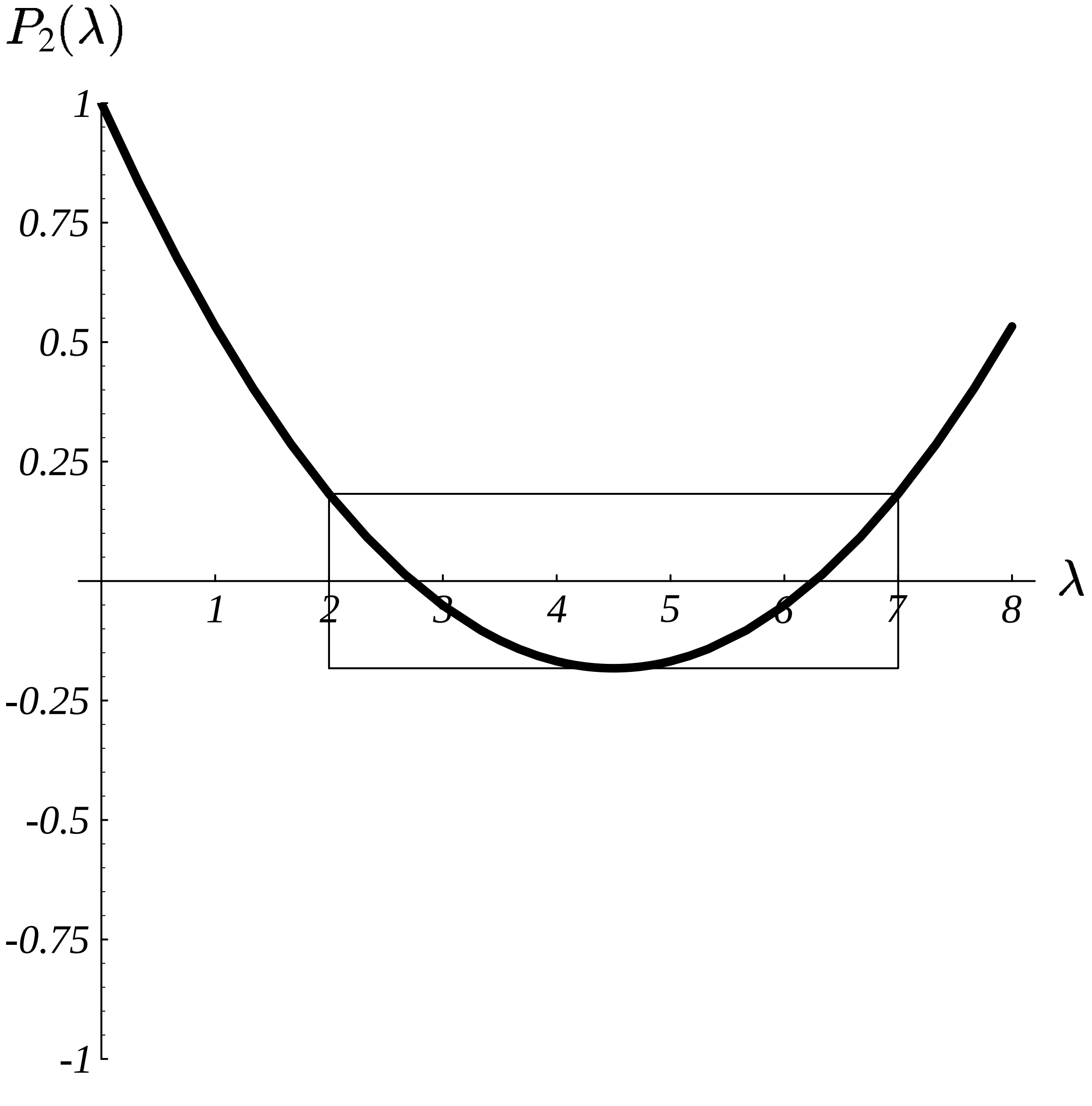
After two steps, CG finds $e_{(2)}$, the point on $\mathcal{E}$ which minimizes $\|e\|_A$.



The convergence of CG after i iterations depends on how close a polynomial P_i of degree i can be to zero on each eigenvalue, given the constraint that $P_i(0) = 1$.

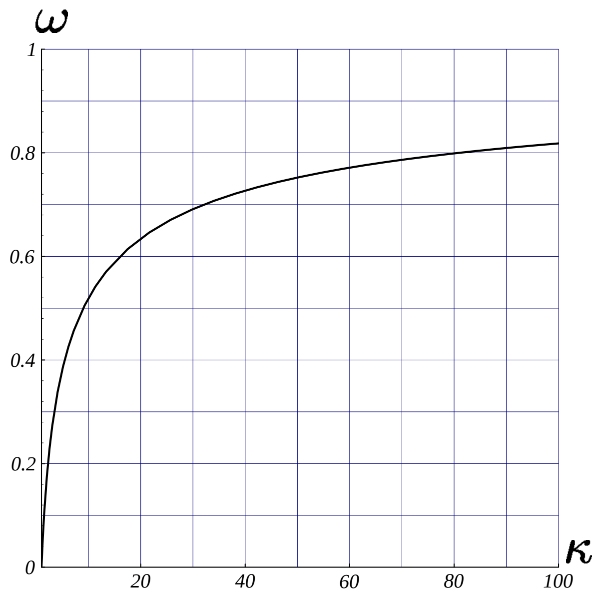


Chebyshev polynomials of degree 2, 5, 10, and 49.

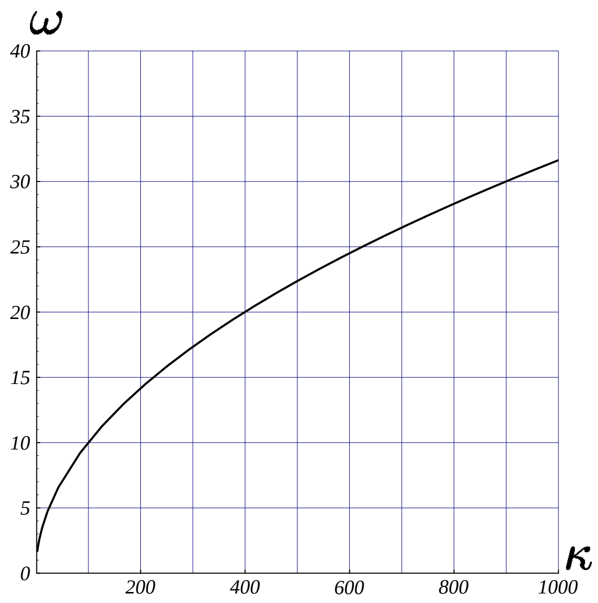


The optimal polynomial $P_2(\lambda)$ for $\lambda_{min} = 2$ and $\lambda_{max} = 7$ in the general case.

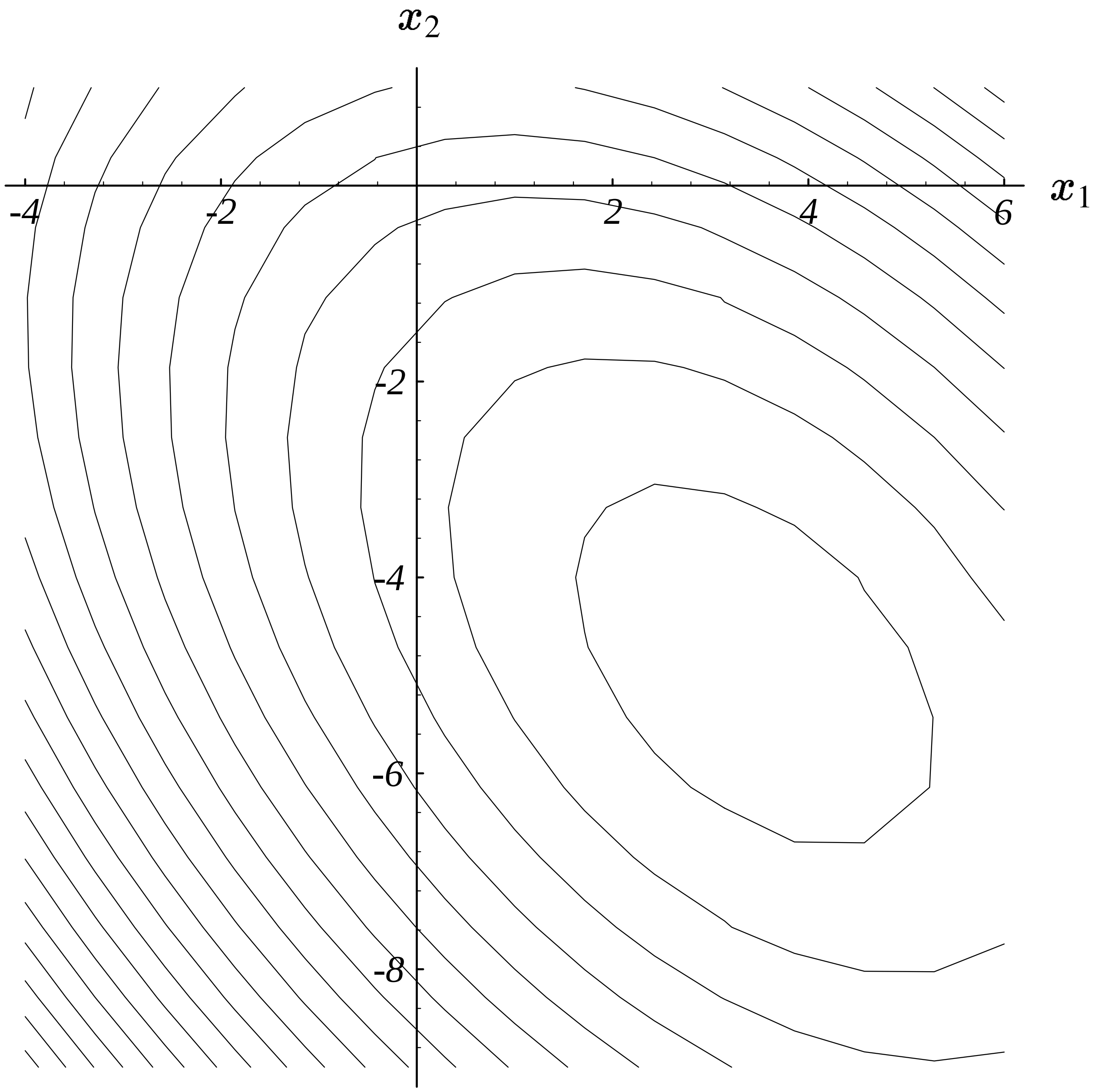
$||e||_A$ is reduced by a factor of at least 0.183 after two iterations of CG.



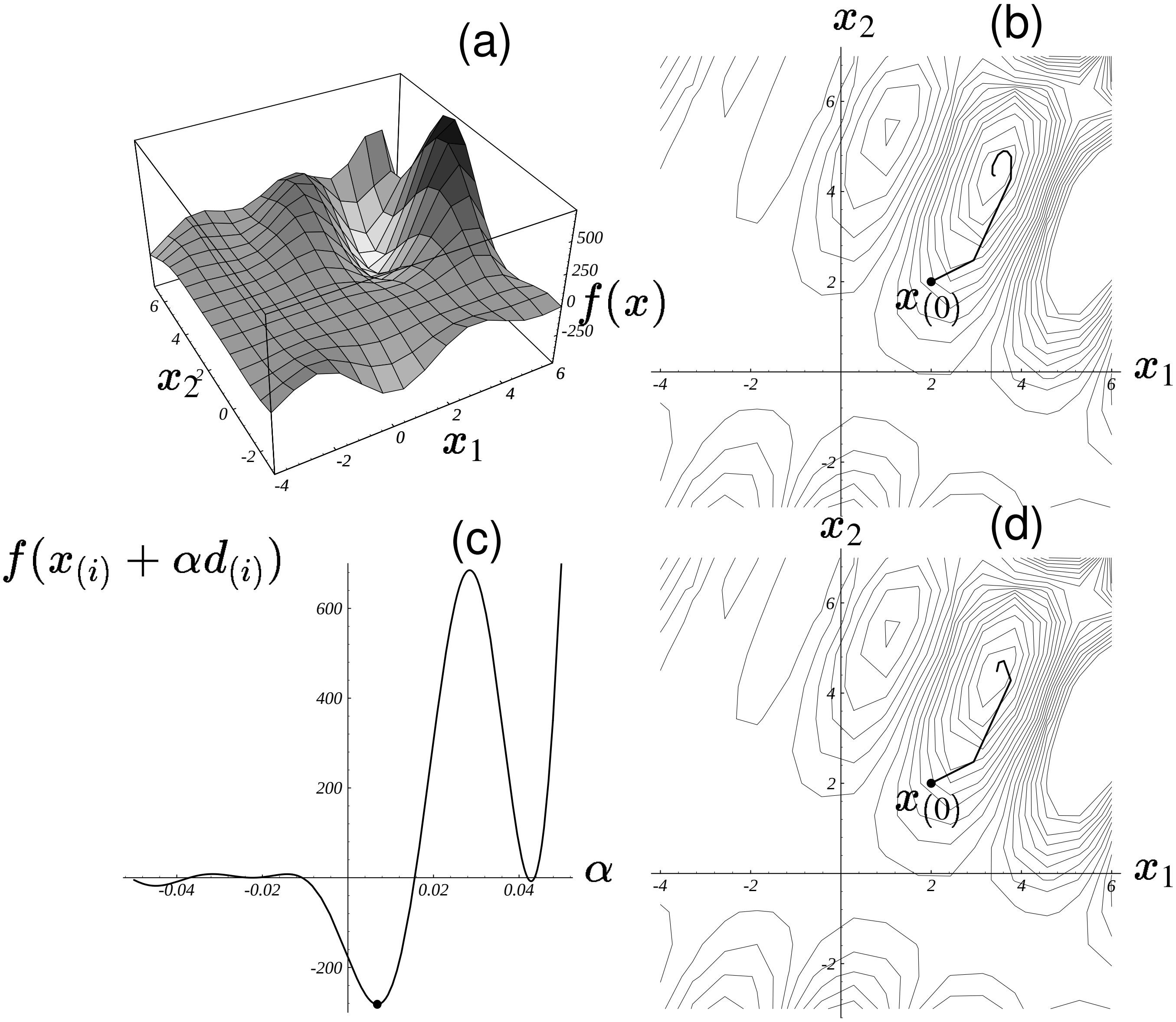
Convergence of Conjugate Gradients (per iteration) as a function of condition number.



Number of iterations of Steepest Descent required to match one iteration of CG.



Contour lines of the quadratic form of the diagonally preconditioned sample problem. The condition number has improved from 3.5 to roughly 2.8.

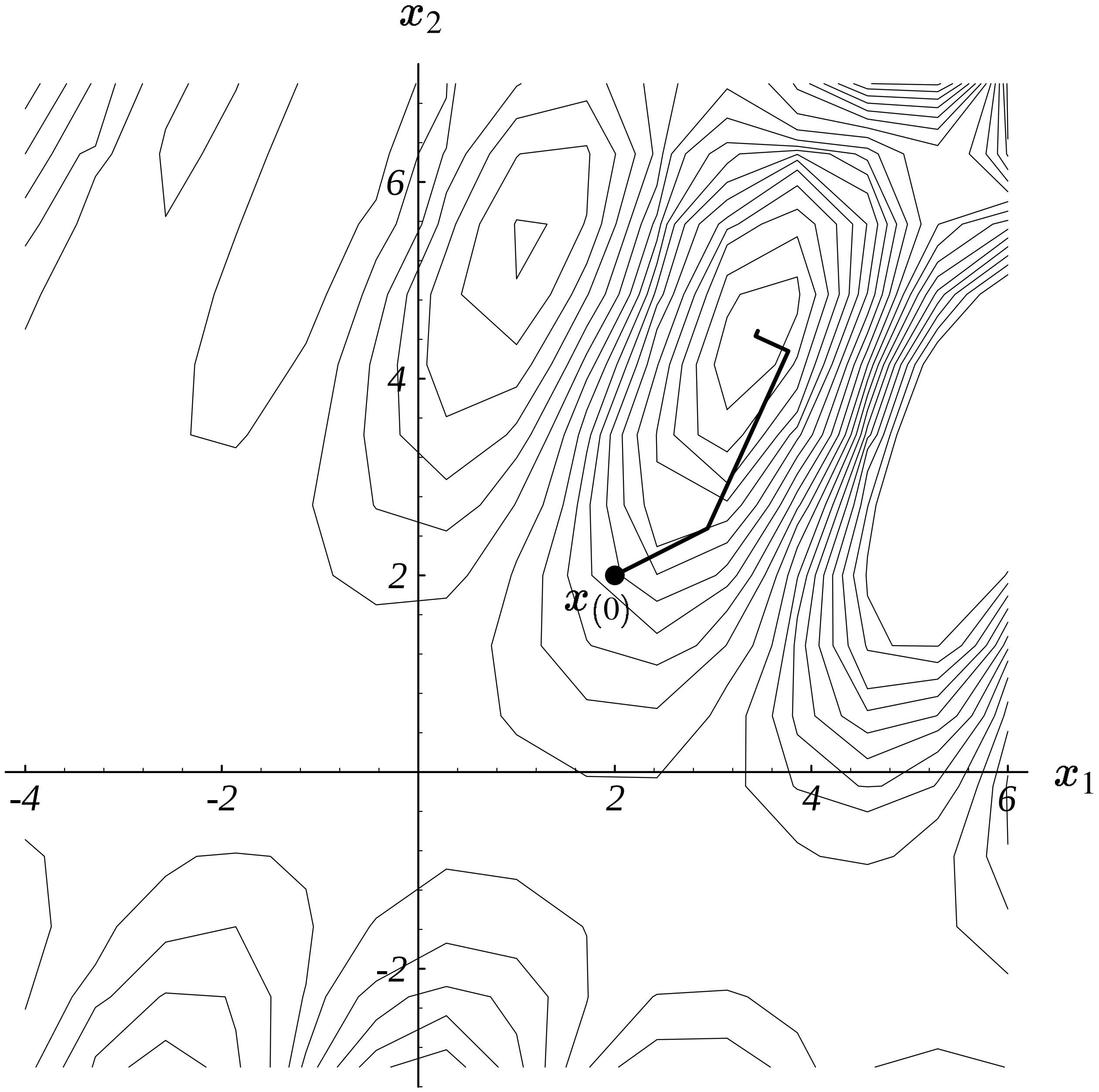


The nonlinear Conjugate Gradient Method.

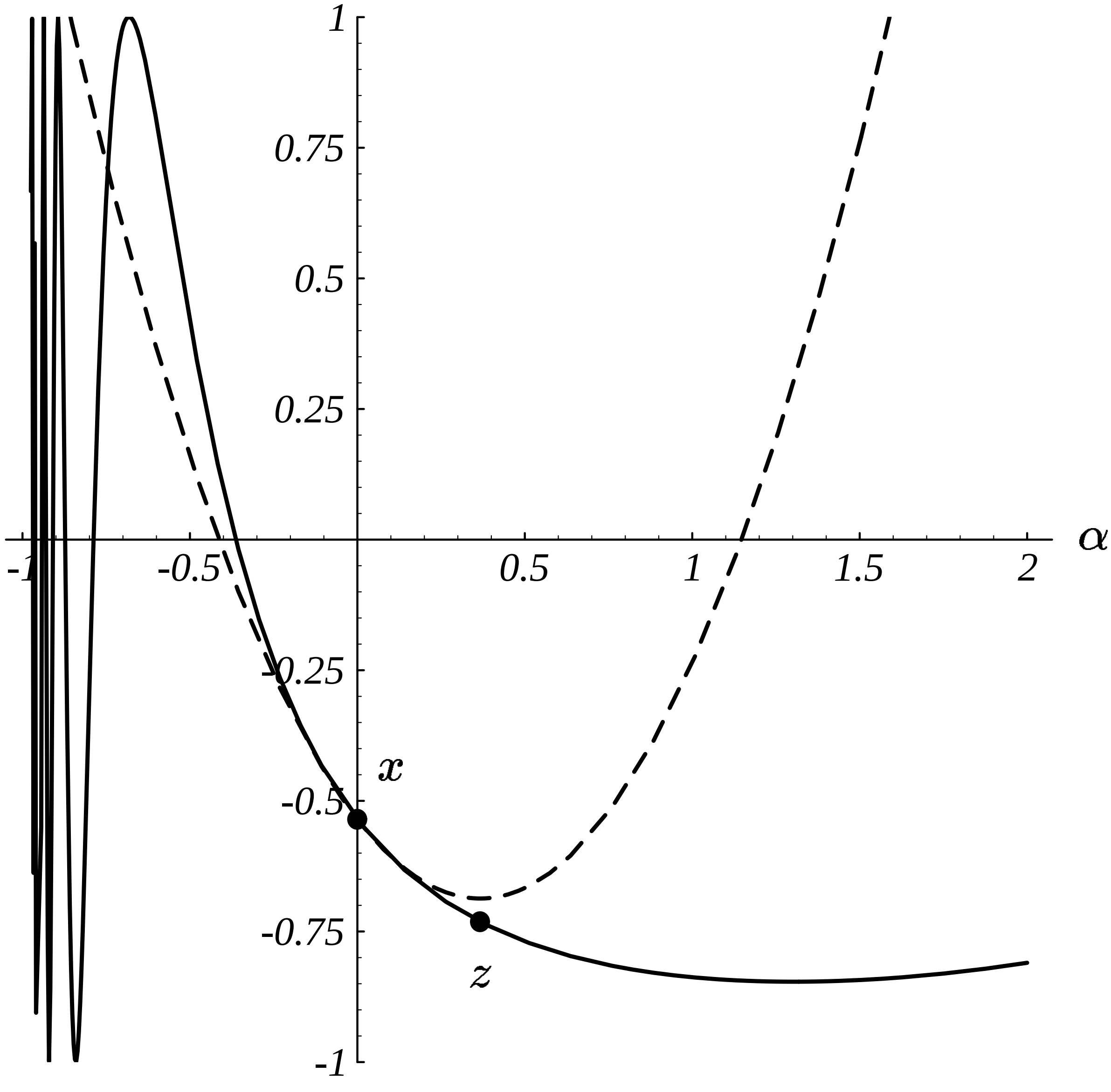
(b) Fletcher-Reeves CG.

(c) Cross-section of the surface corresponding to

(d) Polak-Ribière CG.



Nonlinear CG can be more effective with periodic restarts.

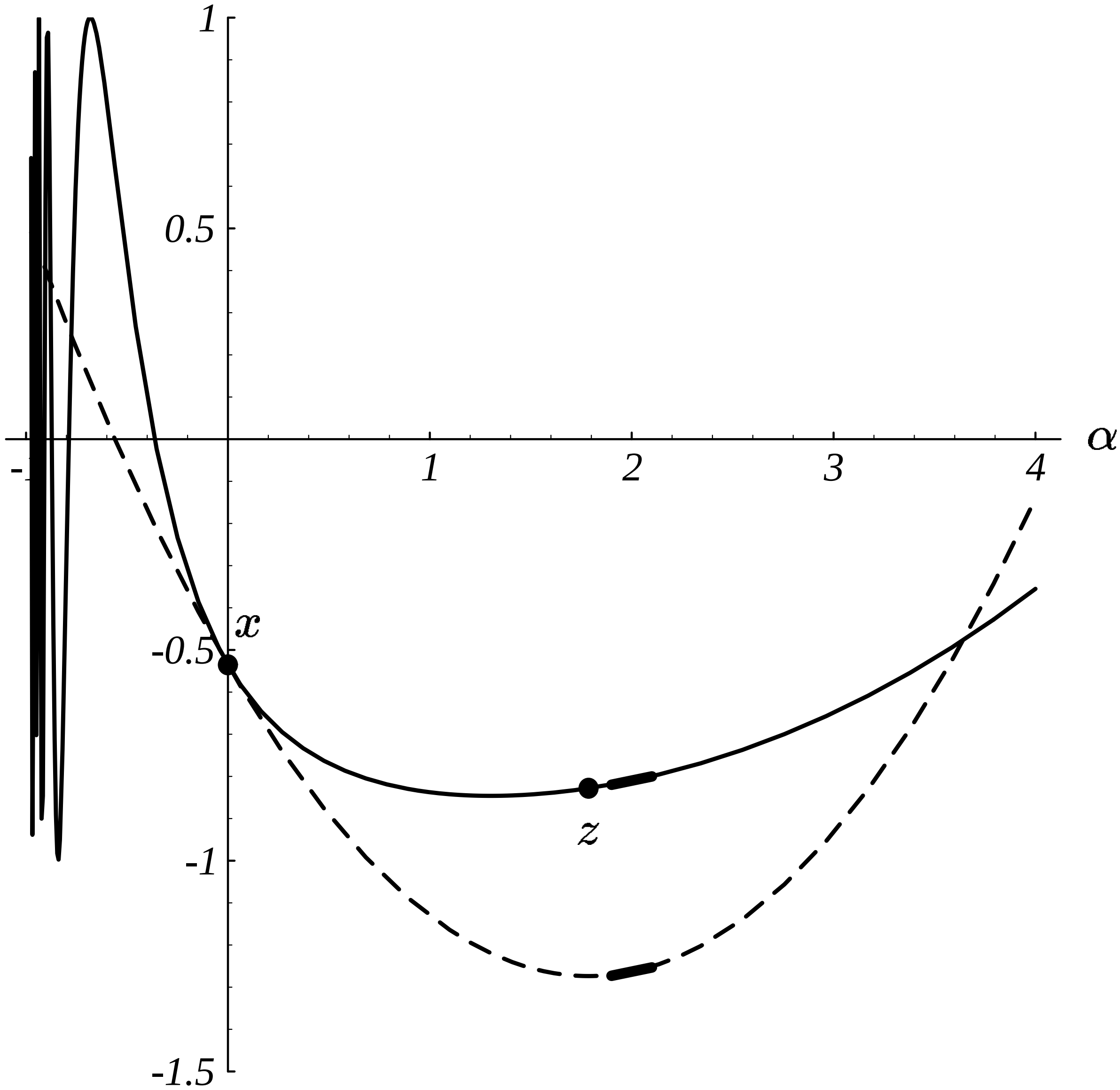


The Newton-Raphson method.

Solid curve: The function to minimize.

Dashed curve: Parabolic approximation to the function, based on first and second derivatives at x.

z is chosen at the base of the parabola.

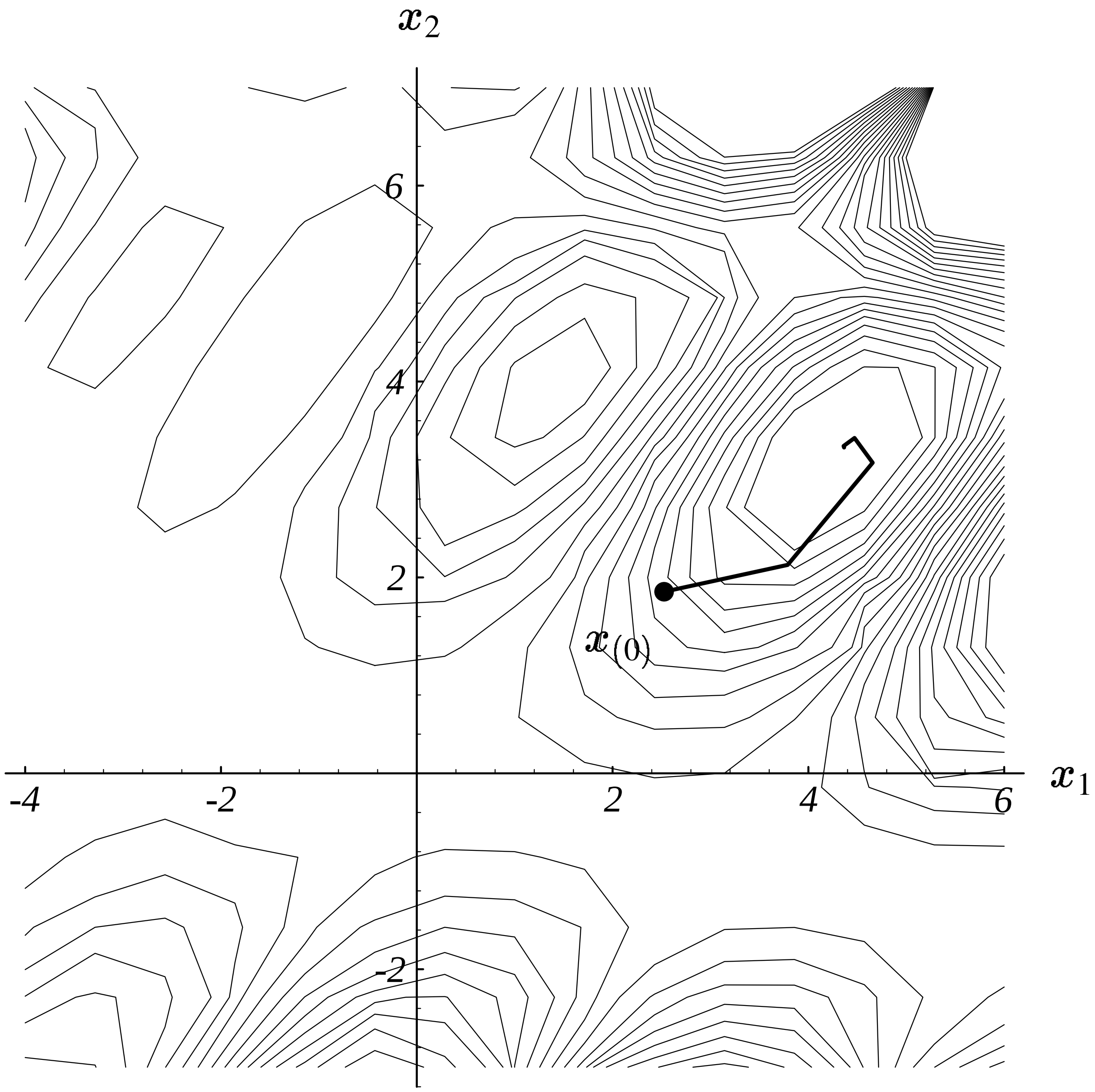


The Secant method.

Solid curve: The function to minimize.

Dashed curve: Parabolic approximation to the function, based on first derivatives at $\alpha = 0$ and $\alpha = 2$.

z is chosen at the base of the parabola.



The preconditioned nonlinear Conjugate Gradient Method.

Polak-Ribière formula and a diagonal preconditioner.

The space has been "stretched" to show the improvement in circularity of the contour lines around the minimum.